CHAPTER 2

Section 2-1

Provide a reasonable description of the sample space for each of the random experiments in Exercises 2-1 to 2-17. There can be more than one acceptable interpretation of each experiment. Describe any assumptions you make.

2-1. Each of three machined parts is classified as either above or below the target specification for the part. Let *a* and *b* denote a part above and below the specification, respectively.

$$S = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$$

2-2. Each of four transmitted bits is classified as either in error or not in error.

Let *e* and *o* denote a bit in error and not in error (*o* denotes okay), respectively.

$$S = \left\{\begin{matrix} eeee, eoee, oeee, ooee, \\ eeeo, eoeo, oeeo, ooeo, \\ eeoe, eooe, oeoe, oooe, \\ eeoo, eooo, oeoo, oooo \end{matrix}\right\}$$

2-3. In the final inspection of electronic power supplies, either units pass, or three types of nonconformities might occur: functional, minor, or cosmetic. Three units are inspected.

Let *a* denote an acceptable power supply.
Let *f*, *m*, and *c* denote a power supply that has a functional, minor, or cosmetic error, respectively.

$$S = \{a, f, m, c\}$$

2-4. The number of hits (views) is recorded at a high-volume Web site in a day.

$S = \{0,1,2,...\}$ = set of nonnegative integers

2-5. Each of 24 Web sites is classified as containing or not containing banner ads.

Let *y* and *n* denote a web site that contains and does not contain banner ads.
The sample space is the set of all possible sequences of *y* and *n* of length 24. An example outcome in the sample space is $S = \{yynnynyyynnynynnnnyynnyy\}$

2-6. An ammeter that displays three digits is used to measure current in milliamperes.

A vector with three components can describe the three digits of the ammeter. Each digit can be 0,1,2,...,9.
The sample space *S* is 1000 possible three digit integers, $S = \{000, 001, ..., 999\}$

2-7. A scale that displays two decimal places is used to measure material feeds in a chemical plant in tons.

S is the sample space of 100 possible two digit integers.

2-8. The following two questions appear on an employee survey questionnaire. Each answer is chosen from the five point scale 1 (never), 2, 3, 4, 5 (always).

Is the corporation willing to listen to and fairly evaluate new ideas?
How often are my coworkers important in my overall job performance?

Let an ordered pair of numbers, such as 43 denote the response on the first and second question. Then, S consists of the 25 ordered pairs $\{11, 12, ..., 55\}$

2-9. The concentration of ozone to the nearest part per billion.

$S = \{0,1,2,...,1E09\}$ in ppb.

2-10. The time until a service transaction is requested of a computer to the nearest millisecond.

$S = \{0,1,2,...,\}$ in milliseconds

2-11. The pH reading of a water sample to the nearest tenth of a unit.

$S = \{1.0,1.1,1.2,\ldots 14.0\}$

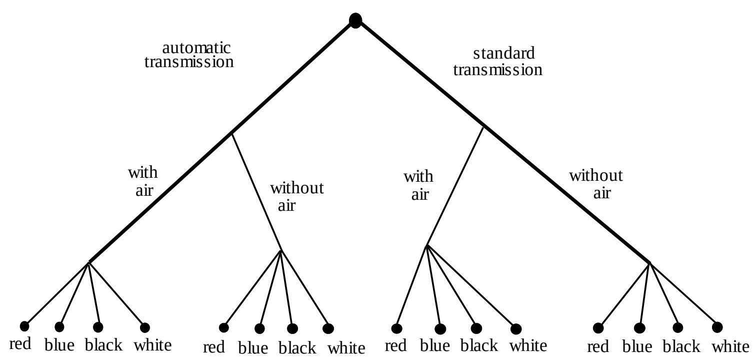
2-12. The voids in a ferrite slab are classified as small, medium, or large. The number of voids in each category is measured by an optical inspection of a sample.

Let *s*, *m*, and *l* denote small, medium, and large, respectively. Then *S* = {*s*, *m*, *l*, *ss*, *sm*, *sl*, ….}

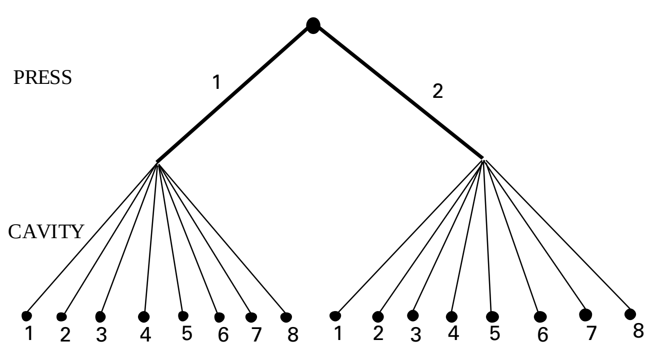
2-13 The time of a chemical reaction is recorded to the nearest millisecond.

$S = \{0,1,2,...,\}$ in milliseconds.

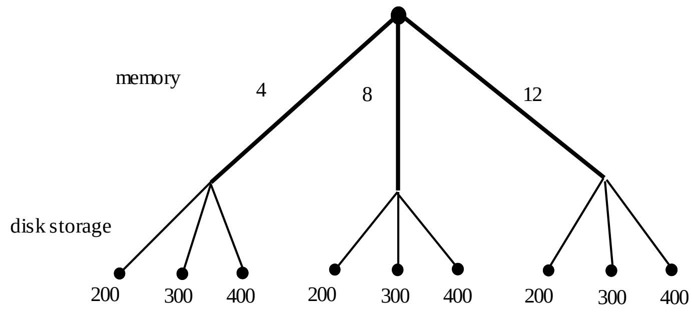
2-14. An order for an automobile can specify either an automatic or a standard transmission, either with or without air conditioning, and with any one of the four colors red, blue, black, or white. Describe the set of possible orders for this experiment.



2-15. A sampled injection-molded part could have been produced in either one of two presses and in any one of the eight cavities in each press.



2-16. An order for a computer system can specify memory of 4, 8, or 12 gigabytes and disk storage of 200, 300, or 400 gigabytes. Describe the set of possible orders.



2-17. Calls are repeatedly placed to a busy phone line until a connection is achieved.

Let c and b denote connect and busy, respectively. Then $S = \{c, bc, bbc, bbbc, bbbbc, \ldots\}$

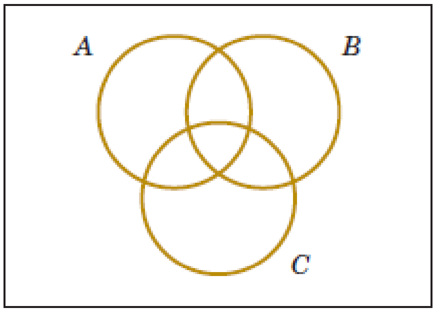
2-18. Three attempts are made to read data in a magnetic storage device before an error recovery procedure that repositions the magnetic head is used. The error recovery procedure attempts three repositionings before an "abort'' message is sent to the operator. Let

s denote the success of a read operation
f denote the failure of a read operation
S denote the success of an error recovery procedure
F denote the failure of an error recovery procedure
A denote an abort message sent to the operator

Describe the sample space of this experiment with a tree diagram.

$$S = \{s, fs, ffs, fffS, fffFS, fffFFS, fffFFFA\}$$

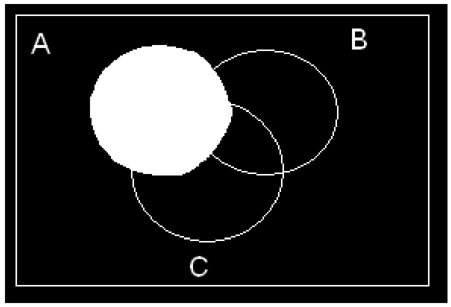
2-19. Three events are shown on the Venn diagram in the following figure:



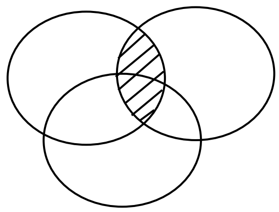
Reproduce the figure and shade the region that corresponds to each of the following events.

(a) A' (b) $A \cap B$ (c) $(A \cap B) \cup C$ (d) $(B \cup C)'$ (e) $(A \cap B)' \cup C$

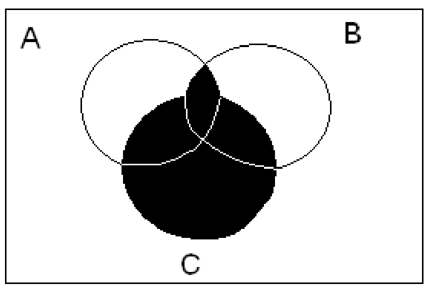
(a)



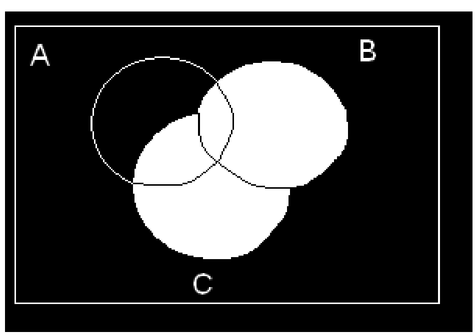
(b)



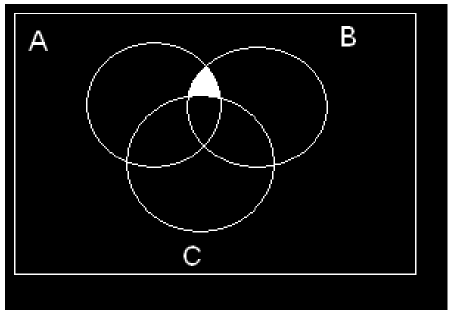
(c)



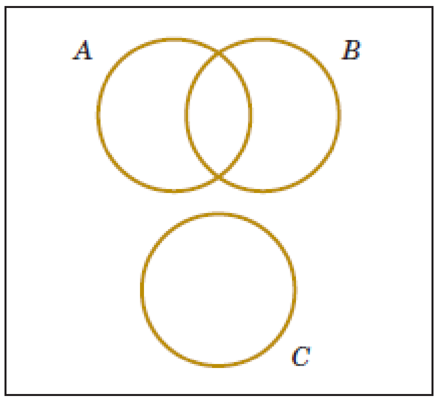
(d)



(e)



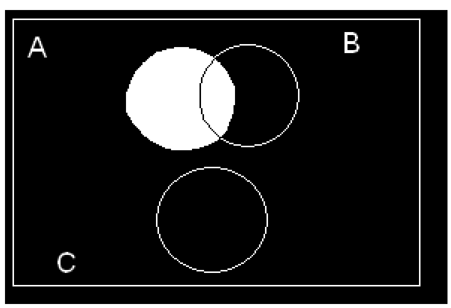
2-20. Three events are shown on the Venn diagram in the following figure:



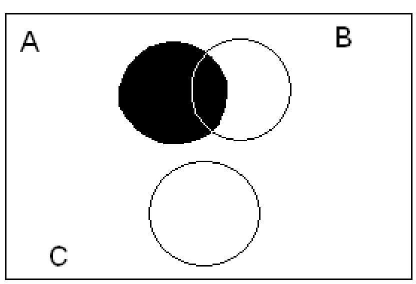
Reproduce the figure and shade the region that corresponds to each of the following events.

(a) A' (b) $(A \cap B) \cup (A \cap B')$ (c) $(A \cap B) \cup C$ (d) $(B \cup C)'$ (e) $(A \cap B)' \cup C$

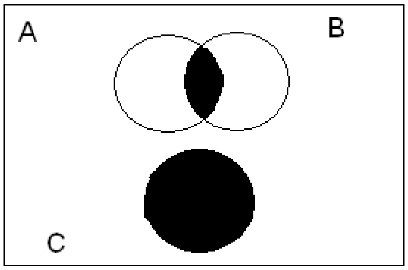
(a)



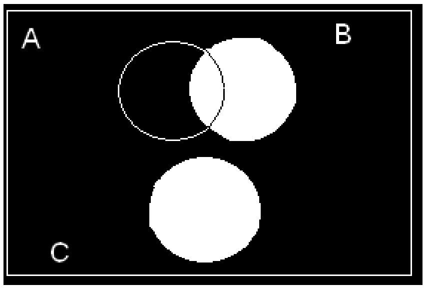
(b)



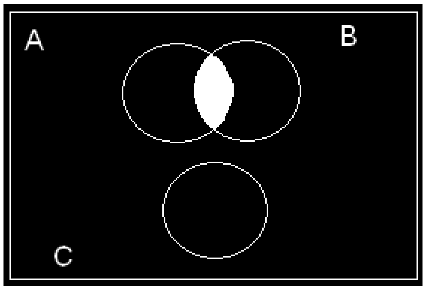
(c)



(d)



(e)



2-21. A digital scale that provides weights to the nearest gram is used.

(a) What is the sample space for this experiment?
Let A denote the event that a weight exceeds 11 grams, let B denote the event that a weight is less than or equal to 15 grams, and let C denote the event that a weight is greater than or equal to 8 grams and less than 12 grams.

Describe the following events.

(b) $A \cup B$ (c) $A \cap B$ (d) A' (e) $A \cup B \cup C$ (f) $(A \cup C)'$ (g) $A \cap B \cap C$ (h) $B' \cap C$ (i) $A \cup (B \cap C)$

(a) Let S = the nonnegative integers from 0 to the largest integer that can be displayed by the scale.
Let X denote the weight.

A is the event that $X > 11$ B is the event that $X \leq 15$ C is the event that $8 \leq X < 12$
$S = \{0, 1, 2, 3, \ldots\}$
(b) S
(c) $11 < X \leq 15$ or $\{12, 13, 14, 15\}$
(d) $X \leq 11$ or $\{0, 1, 2, \ldots, 11\}$
(e) S
(f) $A \cup C$ contains the values of X such that: $X \geq 8$
Thus $(A \cup C)'$ contains the values of X such that: $X < 8$ or $\{0, 1, 2, \ldots, 7\}$
(g) $\varnothing$

(h) B' contains the values of X such that $X > 15$. Therefore, $B' \cap C$ is the empty set. They have no outcomes in common or $\varnothing$.
(i) $B \cap C$ is the event $8 \le X < 12$. Therefore, $A \cup (B \cap C)$ is the event $X \ge 8$ or $\{8, 9, 10, \ldots\}$

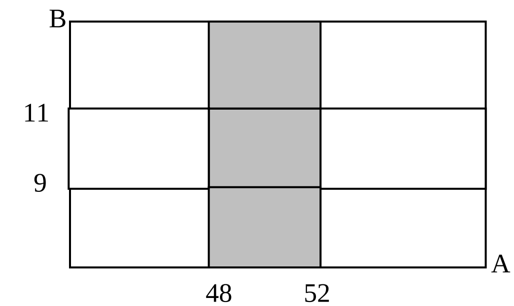
2-22. In an injection-molding operation, the length and width, denoted as X and Y, respectively, of each molded part are evaluated. Let

A denote the event of $48 < X < 52$ centimeters
B denote the event of $9 < Y < 11$ centimeters

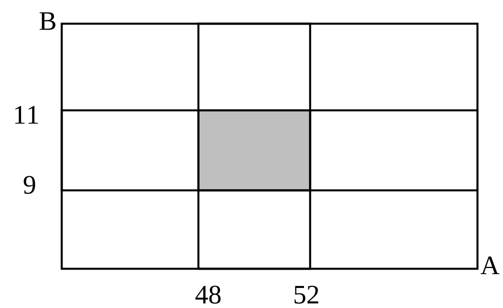
Construct a Venn diagram that includes these events. Shade the areas that represent the following:
(a) A (b) $A \cap B$ (c) $A' \cup B$ (d) $A \cap B$

(e) If these events were mutually exclusive, how successful would this production operation be? Would the process produce parts with $X = 50$ centimeters and $Y = 10$ centimeters?

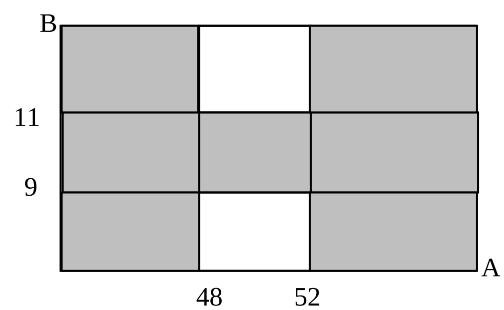
(a)



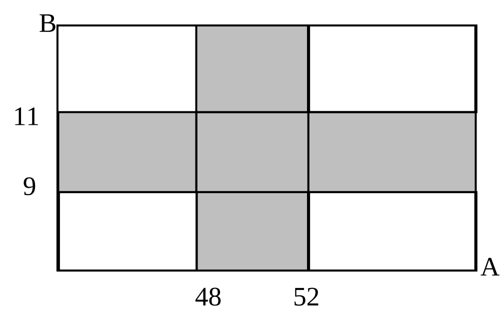
(b)



(c)



(d)



(e) If the events are mutually exclusive, then A∩B is the null set. Therefore, the process does not produce product parts with X = 50 cm and Y = 10 cm. The process would not be successful.

2-23. Four bits are transmitted over a digital communications channel. Each bit is either distorted or received without distortion. Let Ai denote the event that the ith bit is distorted, i = 1,....... ,4.

(a) Describe the sample space for this experiment.
(b) Are the A_i's mutually exclusive?

Describe the outcomes in each of the following events:

(c) A_1 (d) A_1' (e) $A_1 \cap A_2 \cap A_3 \cap A_4$ (f) $(A_1 \cap A_2) \cup (A_3 \cap A_4)$

Let d and o denote a distorted bit and one that is not distorted (o denotes okay), respectively.

(a) $S = \begin{Bmatrix} dddd, dodd, oddd, oodd, \\ dddo, dodo, oddo, oodo, \\ ddod, dood, odod, oood, \\ ddoo, dooo, odoo, oooo \end{Bmatrix}$

(b) No, for example $A_1 \cap A_2 = \{dddd, dddo, ddod, ddoo\}$

(c) $A_1 = \begin{Bmatrix} dddd, dodd, \\ dddo, dodo \\ ddod, dood \\ ddoo, dooo \end{Bmatrix}$

(d) $A_1' = \begin{Bmatrix} oddd, oodd, \\ oddo, oodo, \\ odod, oood, \\ odoo, oooo \end{Bmatrix}$

(e) $A_1 \cap A_2 \cap A_3 \cap A_4 = \{dddd\}$

(f) $(A_1 \cap A_2) \cup (A_3 \cap A_4) = \{dddd, dodd, dddo, oddd, ddod, oodd, ddoo\}$

2-24. In light-dependent photosynthesis, light quality refers to the wavelengths of light that are important. The wavelength of a sample of photosynthetically active radiations (PAR) is measured to the nearest nanometer. The red range is 675–700 nm and the blue range is 450–500 nm. Let A denote the event that PAR occurs in the red range, and let B denote the event that PAR occurs in the blue range. Describe the sample space and indicate each of the following events:

(a) A (b) B (c) $A \cap B$ (d) $A \cup B$

Let w denote the wavelength. The sample space is $\{w \mid w = 0, 1, 2, \ldots\}$
(a) $A=\{w \mid w = 675, 676, \ldots, 700 \text{ nm}\}$
(b) $B=\{ w \mid w = 450, 451, \ldots, 500 \text{ nm}\}$
(c) $A \cap B = \Phi$
(d) $A \cup B = \{w \mid w = 450, 451, \ldots, 500, 675, 676, \ldots, 700 \text{ nm}\}$

2-25. In control replication, cells are replicated over a period of two days. Not until mitosis is completed can freshly synthesized DNA be replicated again. Two control mechanisms have been identified—one positive and one negative. Suppose that a replication is observed in three cells. Let *A* denote the event that all cells are identified as positive, and let *B* denote the event that all cells are negative. Describe the sample space graphically and display each of the following events:

(a) A (b) B (c) $A \cap B$ (d) $A \cup B$

Let *P* and *N* denote positive and negative, respectively.
The sample space is {*PPP*, *PPN*, *PNP*, *NPP*, *PNN*, *NPN*, *NNP*, *NNN*}.

(a) $A = \{ PPP \}$
(b) $B = \{ NNN \}$
(c) $A \cap B = \Phi$
(d) $A \cup B = \{ PPP, NNN \}$

2-26. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized here:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let *A* denote the event that a disk has high shock resistance, and let *B* denote the event that a disk has high scratch resistance. Determine the number of disks in $A \cap B$, A', and $A \cup B$.

$A \cap B = 70$, $A' = 14$, $A \cup B = 95$

2-27. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

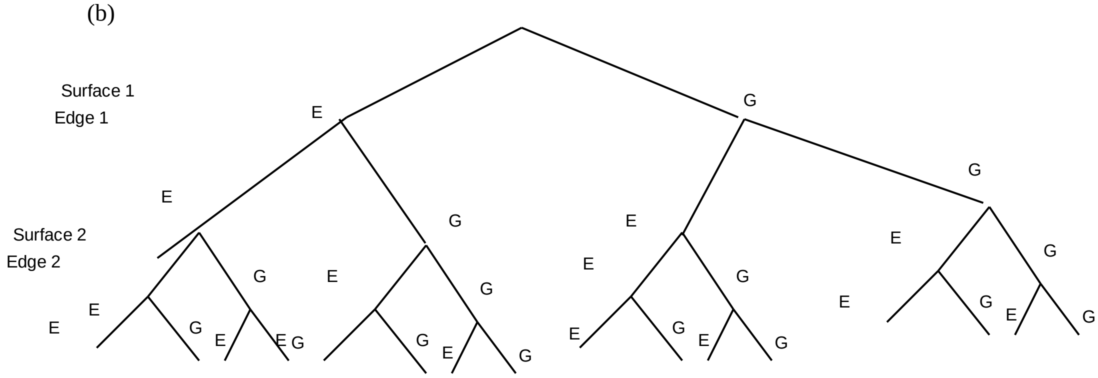
		Edge Finish	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

(a) Let *A* denote the event that a sample has excellent surface finish, and let *B* denote the event that a sample has excellent edge finish. Determine the number of samples in $A' \cap B$, B' and in $A \cup B$.

(b) Assume that each of two samples is to be classified on the basis of surface finish, either excellent or good, and on the basis of edge finish, either excellent or good. Use a tree diagram to represent the possible outcomes of this experiment.

(a) $A' \cap B = 10$, $B' = 10$, $A \cup B = 92$

(b)



2-28. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
	1	22	8
Supplier	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications. Determine the number of samples in $A' \cap B$, B' and in $A \cup B$.

$A' \cap B = 55$, $B' = 23$, $A \cup B = 85$

2-29. The rise time of a reactor is measured in minutes (and fractions of minutes). Let the sample space be positive, real numbers. Define the events A and B as follows:
$A = \{x \mid x < 72.5\}$ and $B = \{x \mid x > 52.5\}$.

Describe each of the following events:

(a) A' (b) B' (c) $A \cap B$ (d) $A \cup B$

(a) $A' = \{x \mid x \geq 72.5\}$
(b) $B' = \{x \mid x \leq 52.5\}$
(c) $A \cap B = \{x \mid 52.5 < x < 72.5\}$
(d) $A \cup B = \{x \mid x > 0\}$

2-30. A sample of two items is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:
(a) The batch contains the items $\{a, b, c, d\}$.
(b) The batch contains the items $\{a, b, c, d, e, f, g\}$.
(c) The batch contains 4 defective items and 20 good items.
(d) The batch contains 1 defective item and 20 good items.

(a) {*ab, ac, ad, bc, bd, cd, ba, ca, da, cb, db, dc*}
(b) {*ab, ac, ad, ae, af, ag, ba, bc, bd, be, bf, bg, ca, cb, cd, ce, cf, cg, da, db, dc, de, df, dg, ea, eb, ec, ed, ef, eg, fa, fb, fc, fg, fd, fe, ga, gb, gc, gd, ge, gf*}, contains 42 elements
(c) Let *d* and *g* denote defective and good, respectively. Then *S* = {*gg, gd, dg, dd*}
(d) S = {*gd, dg, gg*}

2-31. A sample of two printed circuit boards is selected without replacement from a batch. Describe the (ordered) sample space for each of the following batches:

(a) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 2 boards with major defects.

(b) The batch contains 90 boards that are not defective, 8 boards with minor defects, and 1 board with major defects.

Let *g* denote a good board, *m* a board with minor defects, and *j* a board with major defects.

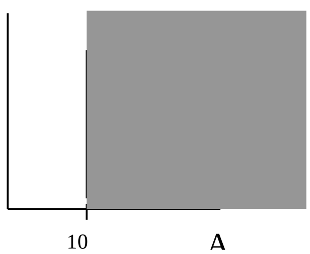
(a) $S = \{gg, gm, gj, mg, mm, mj, jg, jm, jj\}$
(b) $S = \{gg,gm,gj,mg,mm,mj,jg,jm\}$

2-32. Counts of the Web pages provided by each of two computer servers in a selected hour of the day are recorded. Let *A* denote the event that at least 10 pages are provided by server 1, and let *B* denote the event that at least 20 pages are provided by server 2. Describe the sample space for the numbers of pages for the two servers graphically in an *x* *y* plot. Show each of the following events on the sample space graph:

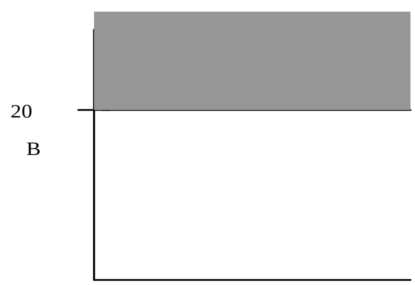
(a) A (b) B (c) $A \cap B$ (d) $A \cup B$

(a) The sample space contains all points in the nonnegative *X*-*Y* plane.

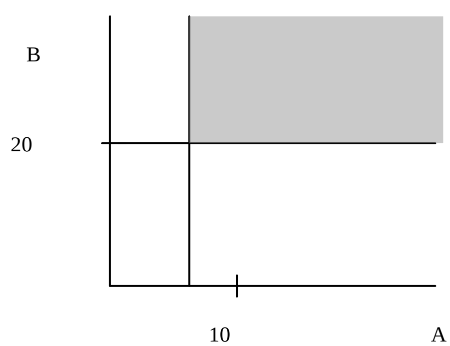
(b)



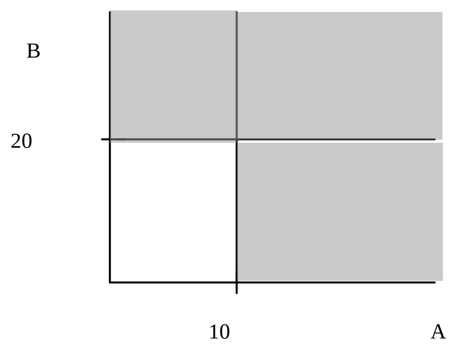
(c)



(d)



(e)

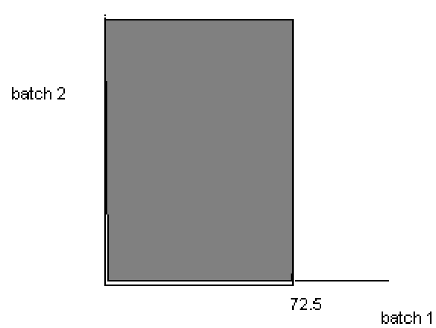


2-33. A reactor's rise time is measured in minutes (and fractions of minutes). Let the sample space for the rise time of each batch be positive, real numbers. Consider the rise times of *two* batches. Let A denote the event that the rise time of batch 1 is less than 72.5 minutes, and let B denote the event that the rise time of batch 2 is greater than 52.5 minutes.

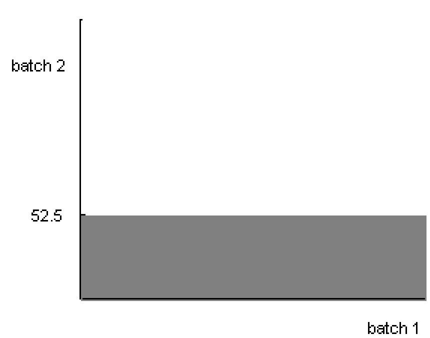
Describe the sample space for the rise time of two batches graphically and show each of the following events on a two dimensional plot:

(a) A (b) B' (c) $A \cap B$ (d) $A \cup B$

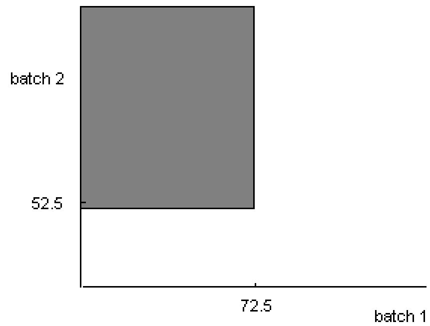
(a)



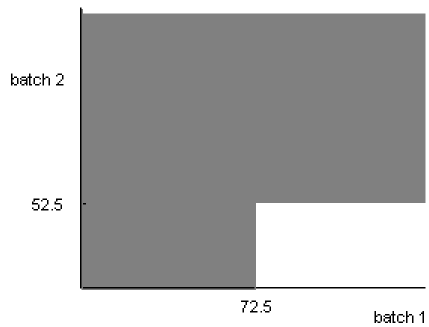
(b)



(c)



(d)



2-34. A wireless garage door opener has a code determined by the up or down setting of 12 switches. How many outcomes are in the sample space of possible codes?

$2^{12} = 4096$

2-35. An order for a computer can specify any one of five memory sizes, any one of three types of displays, and any one of four sizes of a hard disk, and can either include or not include a pen tablet. How many different systems can be ordered?

From the multiplication rule, the answer is $5 \times 3 \times 4 \times 2 = 120$

2-36. In a manufacturing operation, a part is produced by machining, polishing, and painting. If there are three machine tools, four polishing tools, and three painting tools, how many different routings (consisting of machining, followed by polishing, and followed by painting) for a part are possible?

From the multiplication rule, $3 \times 4 \times 3 = 36$

2-37. New designs for a wastewater treatment tank have proposed three possible shapes, four possible sizes, three locations for input valves, and four locations for output valves. How many different product designs are possible?

From the multiplication rule, $3 \times 4 \times 3 \times 4 = 144$

2-38. A manufacturing process consists of 10 operations that can be completed in any order. How many different production sequences are possible?

From equation 2-1, the answer is 10! = 3,628,800

2-39. A manufacturing operation consists of 10 operations. However, five machining operations must be completed before any of the remaining five assembly operations can begin. Within each set of five, operations can be completed in any order. How many different production sequences are possible?

From the multiplication rule and equation 2-1, the answer is 5!5! = 14,400

2-40. In a sheet metal operation, three notches and four bends are required. If the operations can be done in any order, how many different ways of completing the manufacturing are possible?

From equation 2-3, $\frac{7!}{3!4!} = 35$ sequences are possible

2-41. A batch of 140 semiconductor chips is inspected by choosing a sample of 5 chips. Assume 10 of the chips do not conform to customer requirements.

(a) How many different samples are possible?
(b) How many samples of five contain exactly one nonconforming chip?
(c) How many samples of five contain at least one nonconforming chip?

(a) From equation 2-4, the number of samples of size five is $\binom{140}{5} = \frac{140!}{5!135!} = 416{,}965{,}528$

(b) There are 10 ways of selecting one nonconforming chip and there are $\binom{130}{4} = \frac{130!}{4!126!} = 11{,}358{,}880$ ways of selecting four conforming chips. Therefore, the number of samples that contain exactly one nonconforming chip is $10 \times \binom{130}{4} = 113{,}588{,}800$

(c) The number of samples that contain at least one nonconforming chip is the total number of samples $\binom{140}{5}$ minus the number of samples that contain no nonconforming chips $\binom{130}{5}$. That is

$$\binom{140}{5} - \binom{130}{5} = \frac{140!}{5!135!} - \frac{130!}{5!125!} = 130{,}721{,}752$$

2-42. In the layout of a printed circuit board for an electronic product, 12 different locations can accommodate chips.

(a) If five different types of chips are to be placed on the board, how many different layouts are possible?
(b) If the five chips that are placed on the board are of the same type, how many different layouts are possible?

(a) If the chips are of different types, then every arrangement of 5 locations selected from the 12 results in a different layout. Therefore, $P_5^{12} = \frac{12!}{7!} = 95{,}040$ layouts are possible.

(b) If the chips are of the same type, then every subset of 5 locations chosen from the 12 results in a different layout. Therefore, $\binom{12}{5} = \frac{12!}{5!7!} = 792$ layouts are possible.

2-43. In the laboratory analysis of samples from a chemical process, five samples from the process are analyzed daily. In addition, a control sample is analyzed twice each day to check the calibration of the laboratory instruments.

(a) How many different sequences of process and control samples are possible each day? Assume that the five process samples are considered identical and that the two control samples are considered identical.
(b) How many different sequences of process and control samples are possible if we consider the five process samples to be different and the two control samples to be identical?
(c) For the same situation as part (b), how many sequences are possible if the first test of each day must be a control sample?

(a) $\frac{7!}{2!5!} = 21$ sequences are possible.

(b) $\frac{7!}{1!1!1!1!1!2!} = 2520$ sequences are possible.

(c) 6! = 720 sequences are possible.

2-44. In the design of an electromechanical product, 12 components are to be stacked into a cylindrical casing in a manner that minimizes the impact of shocks. One end of the casing is designated as the bottom and the other end is the top.

(a) If all components are different, how many different designs are possible?
(b) If seven components are identical to one another, but the others are different, how many different designs are possible?
(c) If three components are of one type and identical to one another, and four components are of another type and identical to one another, but the others are different, how many different designs are possible?

(a) Every arrangement selected from the 12 different components comprises a different design.
Therefore, $12! = 479{,}001{,}600$ designs are possible.

(b) 7 components are the same, others are different, $\dfrac{12!}{7!1!1!1!1!1!} = 95040$ designs are possible.

(c) $\dfrac{12!}{3!4!} = 3326400$ designs are possible.

2-45. Consider the design of a communication system.

(a) How many three-digit phone prefixes that are used to represent a particular geographic area (such as an area code) can be created from the digits 0 through 9?
(b) As in part (a), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digit?
(c) How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix?

(a) From the multiplication rule, $10^3 = 1000$ prefixes are possible
(b) From the multiplication rule, $8 \times 2 \times 10 = 160$ are possible
(c) Every arrangement of three digits selected from the 10 digits results in a possible prefix.
$P_3^{10} = \dfrac{10!}{7!} = 720$ prefixes are possible.

2-46. A *byte* is a sequence of eight bits and each bit is either 0 or 1.

(a) How many different bytes are possible?
(b) If the first bit of a byte is a parity check, that is, the first byte is determined from the other seven bits, how many different bytes are possible?

(a) From the multiplication rule, $2^8 = 256$ bytes are possible
(b) From the multiplication rule, $2^7 = 128$ bytes are possible

2-47. In a chemical plant, 24 holding tanks are used for final product storage. Four tanks are selected at random and without replacement. Suppose that six of the tanks contain material in which the viscosity exceeds the customer requirements.

(a) What is the probability that exactly one tank in the sample contains high-viscosity material?
(b) What is the probability that at least one tank in the sample contains high-viscosity material?
(c) In addition to the six tanks with high viscosity levels, four different tanks contain material with high impurities. What is the probability that exactly one tank in the sample contains high-viscosity material and exactly one tank in the sample contains material with high impurities?

(a) The total number of samples possible is $\binom{24}{4} = \dfrac{24!}{4!20!} = 10{,}626.$ The number of samples in which exactly one tank has high viscosity is $\binom{6}{1}\binom{18}{3} = \dfrac{6!}{1!5!} \times \dfrac{18!}{3!15!} = 4896$. Therefore, the probability is

$$\frac{4896}{10626} = 0.461$$

(b) The number of samples that contain no tank with high viscosity is $\binom{18}{4} = \frac{18!}{4!14!} = 3060.$ Therefore, the requested probability is $1 - \frac{3060}{10626} = 0.712$.

(c) The number of samples that meet the requirements is $\binom{6}{1}\binom{4}{1}\binom{14}{2} = \frac{6!}{1!5!} \times \frac{4!}{1!3!} \times \frac{14!}{2!12!} = 2184$.

Therefore, the probability is $\frac{2184}{10626} = 0.206$

2-48. Plastic parts produced by an injection-molding operation are checked for conformance to specifications. Each tool contains 12 cavities in which parts are produced, and these parts fall into a conveyor when the press opens. An inspector chooses 3 parts from among the 12 at random. Two cavities are affected by a temperature malfunction that results in parts that do not conform to specifications.

(a) How many samples contain exactly 1 nonconforming part?
(b) How many samples contain at least 1 nonconforming part?

(a) The total number of samples is $\binom{12}{3} = \frac{12!}{3!\,9!} = 220.$ The number of samples that result in one nonconforming part is $\binom{2}{1}\binom{10}{2} = \frac{2!}{1!1!} \times \frac{10!}{2!8!} = 90.$ Therefore, the requested probability is 90/220 = 0.409.

(b) The number of samples with no nonconforming part is $\binom{10}{3} = \frac{10!}{3!7!} = 120.$ The probability of at least one nonconforming part is $1 - \frac{120}{220} = 0.455$.

2-49. A bin of 50 parts contains 5 that are defective. A sample of 10 parts is selected at random, without replacement. How many samples contain at least four defective parts?

From the 5 defective parts, select 4, and the number of ways to complete this step is 5!/(4!1!) = 5
From the 45 non-defective parts, select 6, and the number of ways to complete this step is 45!/(6!39!) = 8,145,060
Therefore, the number of samples that contain exactly 4 defective parts is 5(8,145,060) = 40,725,300
Similarly, from the 5 defective parts, the number of ways to select 5 is 5!(5!1!) = 1
From the 45 non-defective parts, select 5, and the number of ways to complete this step is 45!/(5!40!) = 1,221,759
Therefore, the number of samples that contain exactly 5 defective parts is
1(1,221,759) = 1,221,759

Therefore, the number of samples that contain at least 4 defective parts is
40,725,300 + 1,221,759 = 41,947,059

2-50. The following table summarizes 204 endothermic reactions involving sodium bicarbonate.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is below target. Determine the number of reactions in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cup B'$ (e) $A' \cap B'$

(a) $A \cap B = 56$
(b) $A' = 36 + 56 = 92$
(c) $A \cup B = 40 + 12 + 16 + 44 + 56 = 168$
(d) $A \cup B' = 40+12+16+44+36=148$
(e) $A' \cap B' = 36$

2-51. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. How many different designs are possible?

Total number of possible designs = $4 \times 3 \times 5 \times 3 \times 5 = 900$

2-52. Consider the hospital emergency department data given below. Let *A* denote the event that a visit is to hospital 1, and let *B* denote the event that a visit results in admittance to any hospital.

	Hospital				
	1	**2**	**3**	**4**	**Total**
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Determine the number of persons in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cup B'$ (e) $A' \cap B'$

(a) $A \cap B = 1277$
(b) $A' = 22252 – 5292 = 16960$
(c) $A \cup B = 1685 + 3733 + 1403 + 2 + 14 + 29 + 46 + 3 = 6915$
(d) $A \cup B' = 195 + 270 + 246 + 242 + 3820 + 5163 + 4728 + 3103 + 1277 = 19044$
(e) $A' \cap B' = 270 + 246 + 242 + 5163 + 4728 + 3103 = 13752$

2-53. An article in *The Journal of Data Science* ["A Statistical Analysis of Well Failures in Baltimore County" (2009, Vol. 7, pp. 111–127)] provided the following table of well failures for different geological formation groups in Baltimore County.

	Wells	
Geological Formation Group	**Failed**	**Total**
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

Let *A* denote the event that the geological formation has more than 1000 wells, and let *B* denote the event that a well failed. Determine the number of wells in each of the following events.

(a) $A \cap B$ (b) A' (c) $A \cup B$ (d) $A \cap B'$ (e) $A' \cap B'$

(a) $A \cap B = 170 + 443 + 60 = 673$
(b) $A' = 28 + 363 + 309 + 933 + 39 = 1672$
(c) $A \cup B = 1685 + 3733 + 1403 + 2 + 14 + 29 + 46 + 3 = 6915$
(d) $A \cup B' = 1685 + (28 – 2) + 3733 + (363 – 14) + (309 – 29) + 1403 + (933 – 46) + (39 – 3) = 8399$
(e) $A' \cap B' = 28 – 2 + 363 – 14 + 306 – 29 + 933 – 46 + 39 – 3 = 1578$

2-54. A hospital operating room needs to schedule three knee surgeries and two hip surgeries in a day. Suppose that an operating room needs to handle three knee, four hip, and five shoulder surgeries.

(a) How many different sequences are possible?
(b) How many different sequences have all hip, knee, and shoulder surgeries scheduled consecutively?
(c) How many different schedules begin and end with a knee surgery?

(a) From the formula for the number of sequences $\frac{12!}{3!4!5!}$ = 27,720 sequences are possible.

(b) Combining all hip surgeries into one single unit, all knee surgeries into one single unit and all shoulder surgeries into one unit, the possible number of sequences of these units = 3! = 6

(c)With two surgeries specified, 10 remain and there are $\frac{10!}{4!5!1!}$ = 1,260 different sequences.

2-55. Consider the bar code code 39 is a common bar code system that consists of narrow and wide bars (black) separated by either wide or narrow spaces (white). Each character contains nine elements (five bars and four spaces). The code for a character starts and ends with a bar (either narrow or wide) and a (white) space appears between each bar. The original specification (since revised) used *exactly* two wide bars and one wide space in each character. For example, if *b* and *B* denote narrow and wide (black) bars, respectively, and *w* and *W* denote narrow and wide (white) spaces, a valid character is *bwBwBWbwb* (the number 6). One code is still held back as a delimiter. For each of the following cases, how many characters can be encoded?

(a) The constraint of exactly two wide bars is replaced with one that requires exactly one wide bar.
(b) The constraint of exactly two wide bars is replaced with one that allows either one or two wide bars.
(c) The constraint of exactly two wide bars is dropped.
(d) The constraints of exactly two wide bars and one wide space are dropped.

(a) The constraint of exactly two wide bars is replaced with one that requires exactly one wide bar.
Focus first on the bars. There are 5!/(4!1!) = 5 permutations of the bars with one wide bar and four narrow bars. As in the example, the number of permutations of the spaces = 4. Therefore, the possible number of codes = 5(4) = 20, and if one is held back as a delimiter, 19 characters can be coded.

(b) The constraint of exactly two wide bars is replaced with one that allows either one or two wide bars.
As in the example, the number of codes with exactly two wide bars = 40. From part (a), the number of codes with exactly one wide bar = 20. Therefore, is the possible codes are 40 + 20 = 60, and if one code is held back as a delimiter, 59 characters can be coded.

(c) The constraint of exactly two wide bars is dropped.
There are 2 choices for each bar (wide or narrow) and 5 bars are used in total. Therefore, the number of possibilities for the bars = 2^5 = 32. As in the example, there are 4 possibilities for the spaces. Therefore, the number of codes is 32(3) = 128, and if one is held back as a delimiter, 127 characters can be coded.

(d) The constraints of exactly 2 wide bars and 1 wide space is dropped.
As in part (c), there are 32 possibilities for the bars, and there are also 2^4 = 16 possibilities for the spaces. Therefore, 32(16) = 512 codes are possible, and if one is held back as a delimiter, 511 characters can be coded.

2-56. A computer system uses passwords that contain exactly eight characters, and each character is 1 of the 26 lowercase letters (*a*–*z*) or 26 uppercase letters (*A*–*Z*) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let *A* and *B* denote the events that consist of passwords with only letters or only integers, respectively. Determine the number of passwords in each of the following events.

(a) Ω (b) *A* (c) $A' \cap B'$
(d) Passwords that contain at least 1 integer
(e) Passwords that contain exactly 1 integer

Let |A| denote the number of elements in the set A.
(a) The number of passwords in Ω is $|\Omega| = 62^8$ (from multiplication rule).
(b) The number of passwords in A is $|A| = 52^8$ (from multiplication rule)
(c) $A' \cap B' = (A \cup B)'$. Also, $|A| = 52^8$ and $|B| = 10^8$ and $A \cap B$ = null. Therefore,
$(A \cup B)' = |\Omega| - |A| - |B| = 62^8 - 52^8 - 10^8 \approx 1.65 \times 10^{14}$
(d) Passwords that contain at least 1 integer = $|\Omega| - |A| = 62^8 - 52^8 \approx 1.65 \times 10^{14}$
(e) Passwords that contain exactly 1 integer. The number of passwords with 7 letters is 52^7. Also, 1 integer is selected

in 10 ways, and can be inserted into 8 positions in the password. Therefore, the solution is $8(10)(52^7) \approx 8.22 \times 10^{13}$

2-57. The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C," [*Gastroenterology* (1996, Vol. 111, no. 5, pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let A denote the event that the patient was treated with ribavirin plus interferon alfa, and let B denote the event that the response was complete. Determine the number of patients in each of the following events.

(a) A (b) $A \cap B$ (c) $A \cup B$ (d) $A' \cap B'$

Let |A| denote the number of elements in the set A.
(a) $|A| = 21$
(b) $|A \cap B| = 16$
(c) $|A \cup B| = A+B - (A \cap B) = 21+22 - 16 = 27$
(d) $|A' \cap B'| = 60 - |A \cup B| = 60 - 27 = 33$

Section 2-2

2-58. Each of the possible five outcomes of a random experiment is equally likely. The sample space is $\{a, b, c, d, e\}$. Let A denote the event $\{a, b\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) P(A) (b) P(B) (c) P(A') (d) $P(A \cup B)$ (e) $P(A \cap B)$

All outcomes are equally likely
(a) P(A) = 2/5
(b) P(B) = 3/5
(c) P(A') = 3/5
(d) $P(A \cup B) = 1$
(e) $P(A \cap B) = P(\varnothing) = 0$

2-59. The sample space of a random experiment is $\{a, b, c, d, e\}$ with probabilities 0.1, 0.1, 0.2, 0.4, and 0.2, respectively. Let A denote the event $\{a, b, c\}$, and let B denote the event $\{c, d, e\}$. Determine the following:

(a) P(A) (b) P(B) (c) P(A') (d) $P(A \cup B)$ (e) $P(A \cap B)$

(a) P(A) = 0.4
(b) P(B) = 0.8
(c) P(A') = 0.6
(d) $P(A \cup B) = 1$
(e) $P(A \cap B) = 0.2$

2-60. Orders for a computer are summarized by the optional features that are requested as follows:

	Proportion of Orders
No optional features	0.3
One optional feature	0.5
More than one optional feature	0.2

(a) What is the probability that an order requests at least one optional feature?
(b) What is the probability that an order does not request more than one optional feature?

(a) 0.5 + 0.2 = 0.7

(b) 0.3 + 0.5 = 0.8

2-61. If the last digit of a weight measurement is equally likely to be any of the digits 0 through 9,

(a) What is the probability that the last digit is 0?
(b) What is the probability that the last digit is greater than or equal to 5?

(a) 1/10
(b) 5/10

2-62. A part selected for testing is equally likely to have been produced on any one of six cutting tools.

(a) What is the sample space?
(b) What is the probability that the part is from tool 1?
(c) What is the probability that the part is from tool 3 or tool 5?
(d) What is the probability that the part is not from tool 4?

(a) S = {1, 2, 3, 4, 5, 6}
(b) 1/6
(c) 2/6
(d) 5/6

2-63. An injection-molded part is equally likely to be obtained from any one of the eight cavities on a mold.

(a) What is the sample space?
(b) What is the probability that a part is from cavity 1 or 2?
(c) What is the probability that a part is from neither cavity 3 nor 4?

(a) S = {1,2,3,4,5,6,7,8}
(b) 2/8
(c) 6/8

2-64. In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Because acids and bases are usually colorless (as are the water and salt produced in the neutralization reaction), pH is measured to monitor the reaction. Suppose that the equivalence point is reached after approximately 100 mL of an NaOH solution has been added (enough to react with all the acetic acid present) but that replicates are equally likely to indicate from 95 to 104 mL to the nearest mL. Assume that volumes are measured to the nearest mL and describe the sample space.

(a) What is the probability that equivalence is indicated at 100 mL?
(b) What is the probability that equivalence is indicated at less than 100 mL?
(c) What is the probability that equivalence is indicated between 98 and 102 mL (inclusive)?

The sample space is {95, 96, 97,…, 103, and 104}.

(a) Because the replicates are equally likely to indicate from 95 to 104 mL, the probability that equivalence is indicated at 100 mL is 0.1.
(b) The event that equivalence is indicated at less than 100 mL is {95, 96, 97, 98, 99}. The probability that the event occurs is 0.5.
(c) The event that equivalence is indicated between 98 and 102 mL is {98, 99, 100, 101, 102}. The probability that the event occurs is 0.5.

2-65. In a NiCd battery, a fully charged cell is composed of nickelic hydroxide. Nickel is an element that has multiple oxidation states and that is usually found in the following states:

Nickel Charge	Proportions Found
0	0.17
+2	0.35
+3	0.33
+4	0.15

(a) What is the probability that a cell has at least one of the positive nickel-charged options?
(b) What is the probability that a cell is *not* composed of a positive nickel charge greater than +3?

The sample space is {0, +2, +3, and +4}.

(a) The event that a cell has at least one of the positive nickel charged options is {+2, +3, and +4}. The probability is 0.35 + 0.33 + 0.15 = 0.83.

(b) The event that a cell is not composed of a positive nickel charge greater than +3 is {0, +2, and +3}. The probability is 0.17 + 0.35 + 0.33 = 0.85.

2-66. A credit card contains 16 digits between 0 and 9. However, only 100 million numbers are valid. If a number is entered randomly, what is the probability that it is a valid number?

Total possible: 10^{16}, but only 10^8 are valid. Therefore, P(valid) = $10^8/10^{16} = 1/10^8$

2-67. Suppose your vehicle is licensed in a state that issues license plates that consist of three digits (between 0 and 9) followed by three letters (between *A* and *Z*). If a license number is selected randomly, what is the probability that yours is the one selected?

3 digits between 0 and 9, so the probability of any three numbers is 1/(10*10*10).
3 letters A to Z, so the probability of any three numbers is 1/(26*26*26). The probability your license plate is chosen is then $(1/10^3)*(1/26^3) = 5.7 \times 10^{-8}$

2-68. A message can follow different paths through servers on a network. The sender's message can go to one of five servers for the first step; each of them can send to five servers at the second step; each of those can send to four servers at the third step; and then the message goes to the recipient's server.

(a) How many paths are possible?
(b) If all paths are equally likely, what is the probability that a message passes through the first of four servers at the third step?

(a) 5*5*4 = 100
(b) (5*5)/100 = 25/100=1/4

2-69. Magnesium alkyls are used as homogenous catalysts in the production of linear low-density polyethylene (LLDPE), which requires a finer magnesium powder to sustain a reaction. Redox reaction experiments using four different amounts of magnesium powder are performed. Each result may or may not be further reduced in a second step using three different magnesium powder amounts. Each of these results may or may not be further reduced in a third step using three different amounts of magnesium powder.

(a) How many experiments are possible?
(b) If all outcomes are equally likely, what is the probability that the best result is obtained from an experiment that uses all three steps?
(c) Does the result in part (b) change if five or six or seven different amounts are used in the first step? Explain.

(a) The number of possible experiments is $4 + 4 \times 3 + 4 \times 3 \times 3 = 52$
(b) There are 36 experiments that use all three steps. The probability the best result uses all three steps is 36/52 = 0.6923.
(c) No, it will not change. With k amounts in the first step the number of experiments is $k + 3k + 9k = 13k$. The number of experiments that complete all three steps is $9k$ out of $13k$. The probability is 9/13 = 0.6923.

2-70. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let *A* denote the event that a disk has high shock resistance, and let *B* denote the event that a disk has high scratch resistance. If a disk is selected at random, determine the following probabilities:

(a) P(A) (b) P(B) (c) P(A') (d) $P(A \cap B)$ (e) $P(A \cup B)$ (f) $P(A' \cup B)$

(a) P(A) = 86/100 = 0.86
(b) P(B) = 79/100 = 0.79
(c) P(A') = 14/100 = 0.14
(d) $P(A\cap B)$ = 70/100 = 0.70
(e) $P(A\cup B)$ = (70+9+16)/100 = 0.95
(f) $P(A'\cup B)$ = (70+9+5)/100 = 0.84

2-71. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
Supplier	1	22	8
	2	25	5
	3	30	10

Let *A* denote the event that a sample is from supplier 1, and let *B* denote the event that a sample conforms to specifications. If a sample is selected at random, determine the following probabilities:

(a) P(A) (b) P(B) (c) P(A') (d) $P(A\cap B)$ (e) $P(A\cup B)$ (f) $P(A'\cup B)$

(a) P(A) = 30/100 = 0.30
(b) P(B) = 77/100 = 0.77
(c) P(A') = 1 – 0.30 = 0.70
(d) $P(A\cap B)$ = 22/100 = 0.22
(e) $P(A\cup B)$ = 85/100 = 0.85
(f) $P(A'\cup B)$ =92/100 = 0.92

2-72. An article in the *Journal of Database Management* ["Experimental Study of a Self-Tuning Algorithm for DBMS Buffer Pools" (2005, Vol. 16, pp. 1–20)] provided the workload used in the TPC-C OLTP (Transaction Processing Performance Council's Version C On-Line Transaction Processing) benchmark, which simulates a typical order entry application.

TABLE • 2E-1 Average Frequencies and Operations in TPC-C

Transaction	Frequency	Selects	Updates	Inserts	Deletes	Nonunique Selects	Joins
New order	43	23	11	12	0	0	0
Payment	44	4.2	3	1	0	0.6	0
Order status	4	11.4	0	0	0	0.6	0
Delivery	5	130	120	0	10	0	0
Stock level	4	0	0	0	0	0	1

The frequency of each type of transaction (in the second column) can be used as the percentage of each type of transaction. The average number of *selects* operations required for each type of transaction is shown. Let *A* denote the event of transactions with an average number of *selects* operations of 12 or fewer. Let *B* denote the event of transactions with an average number of *updates* operations of 12 or fewer. Calculate the following probabilities.

(a) P(A) (b) P(B) (c) $P(A\cap B)$ (d) $P(A\cap B')$ (f) $P(A\cup B)$

(a) The total number of transactions is 43+44+4+5+4=100

$$P(A) = \frac{44+4+4}{100} = 0.52$$

(b) $P(B) = \dfrac{100-5}{100} = 0.95$

(c) $P(A\cap B) = \dfrac{44+4+4}{100} = 0.52$

(d) $P(A\cap B') = 0$

(e) $P(A \cup B) = \dfrac{100-5}{100} = 0.95$

2-73. Use the axioms of probability to show the following: $A \cup B$(d) $A \cap B'$

(a) For any event E, $P(E') = 1 - P(E)$. (b) $P(\varnothing) = 0$ (c) If A is contained in B, then $P(A) \le P(B)$.

(a) Because E and E' are mutually exclusive events and $E \cup E'$ = S
1 = P(S) = P($E \cup E'$) = P(E) + P(E'). Therefore, P(E') = 1 - P(E)

(b) Because S and $\varnothing$ are mutually exclusive events with S = $S \cup \varnothing$
P(S) = P(S) + P($\varnothing$). Therefore, P($\varnothing$) = 0

(c) Now, B = $A \cup (A' \cap B)$ and the events A and $A' \cap B$ are mutually exclusive. Therefore,
P(B) = P(A) + P($A' \cap B$). Because P($A' \cap B$) $\geq$ 0 , P(B) $\geq$ P(A).

2.74. Consider the endothermic reaction's table given below. Let *A* denote the event that a reaction's final temperature is 271 K or less. Let *B* denote the event that the heat absorbed is above target.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Determine the following probabilities.

(a) $P(A \cap B)$ (b) $P(A')$ (c) $P(A \cup B)$ (d) $P(A \cup B')$ (e) $P(A' \cap B')$

(a) $P(A \cap B)$ = (40 + 16)/204 = 0.2745
(b) $P(A')$ = (36 + 56)/204 = 0.4510
(c) $P(A \cup B)$ = (40 + 12 + 16 + 44 + 36)/204 = 0.7255
(d) $P(A \cup B')$ = (40 + 12 + 16 + 44 + 56)/204 = 0.8235
(e) $P(A' \cap B')$ = 56/204 = 0.2745

2-75. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. If you visit the site five times, what is the probability that you will not see the same design?

A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. If you visit the site five times, what is the probability that you will not see the same design?

Total number of possible designs is 900. The sample space of all possible designs that may be seen on five visits. This space contains 900^5 outcomes.

The number of outcomes in which all five visits are different can be obtained as follows. On the first visit any one of 900 designs may be seen. On the second visit there are 899 remaining designs. On the third visit there are 898 remaining designs. On the fourth and fifth visits there are 897 and 896 remaining designs, respectively. From the multiplication rule, the number of outcomes where all designs are different is 900*899*898*897*896. Therefore, the probability that a design is not seen again is

$$(900*899*898*897*896)/ 900^5 = 0.9889$$

2-76. Consider the hospital emergency room data is given below. Let *A* denote the event that a visit is to hospital 4, and let *B* denote the event that a visit results in LWBS (at any hospital).

	Hospital				
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Determine the following probabilities.

(a) $P(A \cap B)$ (b) $P(A')$ (c) $P(A \cup B)$ (d) $P(A \cup B')$ (e) $P(A' \cap B')$

(a) $P(A \cap B)$ = 242/22252 = 0.0109
(b) $P(A')$ = (5292+6991+5640)/22252 = 0.8055
(c) $P(A \cup B)$ = (195 + 270 + 246 + 242 + 984 + 3103)/22252 = 0.2265
(d) $P(A \cup B')$ = (4329 + (5292 – 195) + (6991 – 270) + 5640 – 246))/22252 = 0.9680
(e) $P(A' \cap B')$ = (1277 + 1558 + 666 + 3820 + 5163 + 4728)/22252 = 0.7735

2-77. Consider the well failure data is given below. Let *A* denote the event that the geological formation has more than 1000 wells, and let *B* denote the event that a well failed.

	Wells	
Geological Formation Group	Failed	Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

Determine the following probabilities.

(a) $P(A \cap B)$ (b) $P(A')$ (c) $P(A \cup B)$ (d) $P(A \cup B')$ (e) $P(A' \cap B')$

(a) P(A ∩ B) = (170 + 443 + 60)/8493 = 0.0792
(b) P(A′) = (28 + 363 + 309 + 933 + 39)/8493 = 1672/8493 = 0.1969
(c) P(A ∪ B) = (1685+3733+1403+2+14+29+46+3)/8493 = 6915/8493 = 0.8142
(d) P(A ∪ B′) = (1685 + (28 – 2) + 3733 + (363 – 14) + (309 – 29) + 1403 + (933 – 46) + (39 – 3))/8493 = 8399/8493 = 0.9889
(e) P(A′ ∩ B′) = (28 – 2 + 363 – 14 + 306 – 29 + 933 – 46 + 39 – 3)/8493 = 1578/8493 = 0.1858

2-78. Consider the bar code 39 is a common bar code system that consists of narrow and wide bars (black) separated by either wide or narrow spaces (white). Each character contains nine elements (five bars and four spaces). The code for a character starts and ends with a bar (either narrow or wide) and a (white) space appears between each bar. The original specification (since revised) used *exactly* two wide bars and one wide space in each character. For example, if *b* and *B* denote narrow and wide (black) bars, respectively, and *w* and *W* denote narrow and wide (white) spaces, a valid character is *bwBwBWbwb* (the number 6). Suppose that all 40 codes are equally likely (none is held back as a delimiter).

Determine the probability for each of the following:
(a) A wide space occurs before a narrow space.
(b) Two wide bars occur consecutively.
(c) Two consecutive wide bars are at the start or end.
(d) The middle bar is wide.

(a) There are 4 spaces and exactly one is wide.
Number of permutations of the spaces where the wide space appears first is 1.
Number of permutations of the bars is 5!/(2!3!) = 10.
Total number of permutations where a wide space occurs before a narrow space 1(10) = 10.

P(wide space occurs before a narrow space) =10/40 = 1/4

(b) There are 5 bars and 2 are wide.
The spaces are handled as in part (a).
Number of permutations of the bars where 2 wide bars are consecutive is 4.
Therefore, the probability is 16/40 = 0.4

(c) The spaces are handled as in part (a).
Number of permutations of the bars where the 2 consecutive wide bars are at the start or end is 2. Therefore, the probability is 8/40 = 0.2

(d) The spaces are handled as in part (a).
Number of permutations of the bars where a wide bar is at the center is 4 because there are 4 remaining positions for the second wide bar. Therefore, the probability is 16/40 = 0.4.

2-79. A hospital operating room needs to schedule three knee surgeries and two hip surgeries in a day. Suppose that an operating room needs to schedule three knee, four hip, and five shoulder surgeries. Assume that all schedules are equally likely.

Determine the probability for each of the following:
(a) All hip surgeries are completed before another type of surgery.
(b) The schedule begins with a hip surgery.
(c) The fi rst and last surgeries are hip surgeries.
(d) The fi rst two surgeries are hip surgeries.

(a) P(all hip surgeries before another type)= $\frac{\frac{8!}{3!5!}}{\frac{12!}{3!4!5!}} = \frac{8!4!}{12!} = \frac{1}{495} = 0.00202$

(b) P(begins with hip surgery)= $\frac{\frac{11!}{3!3!5!}}{\frac{12!}{3!4!5!}} = \frac{11!4!}{12!3!} = \frac{1}{3}$

(c) P(first and last are hip surgeries)= $\frac{\frac{10!}{2!3!5!}}{\frac{12!}{3!4!5!}} = \frac{1}{11}$

(d) P(first two are hip surgeries)= $\frac{\frac{10!}{2!3!5!}}{\frac{12!}{3!4!5!}} = \frac{1}{11}$

2-80. Suppose that a patient is selected randomly from the those described ,The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C," [*Gastroenterology* (1996, Vol. 111, no. 5, pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let *A* denote the event that the patient is in the group treated with interferon alfa, and let *B* denote the event that the patient has a complete response.
Determine the following probabilities.

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$ (e) $P(A' \cup B)$

(a) P(A)= 19/60 = 0.3167
(b) P(B)= 22/60 = 0.3667
(c) P(A∩B) = 6/60 = 0.1
(d) P(AUB) = P(A)+P(B)-P(A∩B) = (19+22-6)/60 =0.5833
(e) P(A'UB) = P(A')+P(B)-P(A'∩B) = $\frac{21+20}{60} + \frac{22}{60} - \frac{16}{60} = 0.7833$

2-81. A computer system uses passwords that contain exactly eight characters, and each character is one of 26 lowercase letters (*a*–*z*) or 26 uppercase letters (*A*–*Z*) or 10 integers (0–9). Let Ω denote the set of all possible passwords, and let *A* and *B* denote the events that consist of passwords with only letters or only integers, respectively. Suppose that all passwords in Ω are equally likely.
Determine the probability of each of the following:
(a) *A* (b) *B*
(c) A password contains at least 1 integer.
(d) A password contains exactly 2 integers.

(a) P(A) = $\frac{52^8}{62^8} = 0.2448$
(b) P(B) = $\frac{10^8}{62^8}$ = $4.58\text{x}10^{-7}$
(c) P(contains at least 1 integer) =1 - P(password contains no integer) = 1 - $\frac{52^8}{62^8}$ = 0.7551
(d) P(contains exactly 2 integers)
Number of positions for the integers is 8!/(2!6!) = 28
Number of permutations of the two integers is 10^2 = 100
Number of permutations of the six letters is 52^6
Total number of permutations is 62^8
Therefore, the probability is
$\frac{28(100)(52^6)}{62^8} = 0.254$

<u>Section 2-3</u>

2-82. If $P(A) = 0.3$, $P(B) = 0.2$, and $P(A \cap B) = 0.1$, determine the following probabilities:

(a) $P(A')$ (b) $P(A \cup B)$ (c) $P(A' \cap B)$ (d) $P(A \cap B')$ (e) $P[(A \cup B)']$ (f) $P(A' \cup B)$

(a) P(A') = 1- P(A) = 0.7
(b) P ($A \cup B$) = P(A) + P(B) - P($A \cap B$) = 0.3+0.2 - 0.1 = 0.4
(c) P($A' \cap B$) + P($A \cap B$) = P(B). Therefore, P($A' \cap B$) = 0.2 - 0.1 = 0.1
(d) P(A) = P($A \cap B$) + P($A \cap B'$). Therefore, P($A \cap B'$) = 0.3 - 0.1 = 0.2
(e) P(($A \cup B$)') = 1 - P($A \cup B$) = 1 - 0.4 = 0.6
(f) P($A' \cup B$) = P(A') + P(B) - P($A' \cap B$) = 0.7 + 0.2 - 0.1 = 0.8

2-83. If *A*, *B*, and *C* are mutually exclusive events with $P(A) = 0.2$, $P(B) = 0.3$, and $P(C) = 0.4$, determine the following probabilities:

(a) $P(A \cup B \cup C)$ (b) $P(A \cap B \cap C)$ (c) $P(A \cap B)$ (d) $P[(A \cup B) \cap C]$ (e) $P(A' \cap B' \cap C')$

(a) P($A \cup B \cup C$) = P(A) + P(B) + P(C), because the events are mutually exclusive. Therefore,
P($A \cup B \cup C$) = 0.2+0.3+0.4 = 0.9
(b) P ($A \cap B \cap C$) = 0, because $A \cap B \cap C = \varnothing$
(c) P($A \cap B$) = 0 , because $A \cap B = \varnothing$
(d) P($(A \cup B) \cap C$) = 0, because $(A \cup B) \cap C = (A \cap C) \cup (B \cap C) = \varnothing$
(e) P($A' \cap B' \cap C'$) =1-[P(A) + P(B) + P(C)] = 1-(0.2+0.3+0.4) = 0.1

2-84. In the article “ACL Reconstruction Using Bone-Patellar Tendon-Bone Press-Fit Fixation: 10-Year Clinical Results” in *Knee Surgery, Sports Traumatology, Arthroscopy* (2005, Vol. 13, pp. 248–255), the following causes for knee injuries were considered:

Activity	Percentage of Knee Injuries
Contact sport	46%
Noncontact sport	44%
Activity of daily living	9%
Riding motorcycle	1%

(a) What is the probability that a knee injury resulted from a sport (contact or noncontact)?
(b) What is the probability that a knee injury resulted from an activity other than a sport?

(a) P(Caused by sports) = P(Caused by contact sports or by noncontact sports)
= P(Caused by contact sports) + P(Caused by noncontact sports)
= 0.46 + 0.44 = 0.9
(b) 1- P(Caused by sports) = 0.1

2.85. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

(a) If a disk is selected at random, what is the probability that its scratch resistance is high and its shock resistance is high?
(b) If a disk is selected at random, what is the probability that its scratch resistance is high or its shock resistance is high?
(c) Consider the event that a disk has high scratch resistance and the event that a disk has high shock resistance. Are these two events mutually exclusive?

(a) 70/100 = 0.70
(b) (79+86-70)/100 = 0.95
(c) No, $P(A \cap B) \neq 0$

2-86. Strands of copper wire from a manufacturer are analyzed for strength and conductivity. The results from 100 strands are as follows:

	Strength	
	High	Low
High conductivity	74	8
Low conductivity	15	3

(a) If a strand is randomly selected, what is the probability that its conductivity is high and its strength is high?
(b) If a strand is randomly selected, what is the probability that its conductivity is low or its strength is low?
(c) Consider the event that a strand has low conductivity and the event that the strand has low strength. Are these two events mutually exclusive?

(a) P(High temperature and high conductivity)= 74/100 =0.74
(b) P(Low temperature or low conductivity)
= P(Low temperature) + P(Low conductivity) – P(Low temperature and low conductivity)
= (8+3)/100 + (15+3)/100 – 3/100
= 0.26
(c) No, they are not mutually exclusive. Because P(Low temperature) + P(Low conductivity)
= (8+3)/100 + (15+3)/100
= 0.29, which is not equal to P(Low temperature or low conductivity).

2-87. The analysis of shafts for a compressor is summarized by conformance to specifications.

		Roundness Conforms	
		Yes	No
Surface Finish	Yes	345	5
Conforms	No	12	8

(a) If a shaft is selected at random, what is the probability that it conforms to surface finish requirements?
(b) What is the probability that the selected shaft conforms to surface finish requirements or to roundness requirements?
(c) What is the probability that the selected shaft either conforms to surface finish requirements or does not conform to roundness requirements?
(d) What is the probability that the selected shaft conforms to both surface finish and roundness requirements?

(a) 350/370

(b) $\frac{345+5+12}{370}=\frac{362}{370}$

(c) $\frac{345+5+8}{370}=\frac{358}{370}$

(d) 345/370

2-88. Cooking oil is produced in two main varieties: mono and polyunsaturated. Two common sources of cooking oil are corn and canola. The following table shows the number of bottles of these oils at a supermarket:

		Type of oil	
		Canola	Corn
Type of Unsaturation	Mono	7	13
	Poly	93	77

(a) If a bottle of oil is selected at random, what is the probability that it belongs to the polyunsaturated category?
(b) What is the probability that the chosen bottle is monounsaturated canola oil?

(a) 170/190 = 17/19
(b) 7/190

2-89. A manufacturer of front lights for automobiles tests lamps under a high-humidity, high-temperature environment using intensity and useful life as the responses of interest. The following table shows the performance of 130 lamps:

		Useful life	
		Satisfactory	Unsatisfactory
Intensity	Satisfactory	117	3
	Unsatisfactory	8	2

(a) Find the probability that a randomly selected lamp will yield unsatisfactory results under any criteria.
(b) The customers for these lamps demand 95% satisfactory results. Can the lamp manufacturer meet this demand?

(a) P(unsatisfactory) = (5 + 10 – 2)/130 = 13/130
(b) P(both criteria satisfactory) = 117/130 = 0.90, No

2-90. A computer system uses passwords that are six characters, and each character is one of the 26 letters (*a*–*z*) or 10 integers (0–9). Uppercase letters are not used. Let *A* denote the event that a password begins with a vowel (either *a*, *e*, *i*, *o*, or *u*), and let *B* denote the event that a password ends with an even number (either 0, 2, 4, 6, or 8). Suppose a hacker selects a password at random. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

(a) 5/36
(b) 5/36
(c) $P(A \cap B) = P(A)P(B) = 25/1296$
(d) $P(A \cup B) = P(A) + P(B) - P(A)P(B) = 10/36 - 25/1296 = 0.2585$

2-91. Consider the endothermic reactions given below. Let *A* denote the event that a reaction's final temperature is 271 K or less. Let *B* denote the event that the heat absorbed is above target.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cap B')$ (c) $P(A' \cup B')$

$P(A)$ = 112/204 = 0.5490, P(B) = 92/204 = 0.4510, P(A ∩ B) = (40+16)/204 = 0.2745
(a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ = 0.5490 + 0.4510 – 0.2745 = 0.7255
(b) $P(A \cap B')$ = (12 + 44)/204 = 0.2745 and $P(A \cup B') = P(A) + P(B') - P(A \cap B')$ = 0.5490 + (1 – 0.4510) – 0.2745 = 0.8235
(c) $P(A' \cup B') = 1 - P(A \cap B)$ = 1 – 0.2745 = 0.7255

2-92. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let *A* denote the event that the design color is red, and let *B* denote the event that the font size is not the smallest one. Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cup B')$ (c) $P(A' \cup B')$

P(A) = 1/4 = 0.25, P(B) = 4/5 = 0.80, P(A ∩ B) = P(A)P(B) = (1/4)(4/5) = 1/5 = 0.20
(a) P(A ∪ B) = P(A) + P(B) – P(A ∩ B) = 0.25 + 0.80 – 0.20 = 0.85
(b) First P(A ∩ B') = P(A)P(B′) = (1/4)(1/5) = 1/20 = 0.05. Then P(A ∪ B′) = P(A) + P(B′) – P(A ∩ B') = 0.25 + 0.20 – 0.05= 0.40
(c) P(A′ ∪ B′) = 1 – P(A ∩ B) = 1 – 0.20 = 0.80

2-93. Consider the hospital emergency room data given below. Let *A* denote the event that a visit is to hospital 4, and let *B* denote the event that a visit results in LWBS (at any hospital).

	Hospital				
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cup B')$ (c) $P(A' \cup B')$

P(A) = 4329/22252 = 0.1945, P(B) = 953/22252 = 0.0428, P(A ∩ B) = 242/22252 = 0.0109,
P(A ∩ B′) = (984+3103)/22252 = 0.1837
(a) P(A ∪ B) = P(A) + P(B) – P(A ∩ B) = 0.1945 + 0.0428 – 0.0109 = 0.2264
(b) P(A ∪ B′) = P(A) + P(B′) – P(A ∩ B′) = 0.1945 + (1 – 0.0428) – 0.1837 = 0.9680
(c) P(A′ ∪ B′) = 1 – P(A ∩ B) = 1 – 0.0109 = 0.9891

2-94. Consider the well failure data given below. Let *A* denote the event that the geological formation has more than 1000 wells, and let *B* denote the event that a well failed.

	Wells	
Geological Formation Group	Failed	Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

Use the addition rules to calculate the following probabilities.

(a) $P(A \cup B)$ (b) $P(A \cup B')$ (c) $P(A' \cup B')$

P(A) = (1685 + 3733 + 1403)/8493 = 0.8031, P(B) = (170 + 2 + 443 + 14 + 29 + 60 + 46 + 3)/8493 = 0.0903,
P(A ∩ B) = (170 + 443 + 60)/8493 = 0.0792, P(A ∩ B′) = (1515+3290+1343)/8493 = 0.7239
a) P(A ∪ B) = P(A) + P(B) – P(A ∩ B) = 0.8031 + 0.0903 – 0.0792 = 0.8142
b) P(A ∪ B′) = P(A) + P(B′) – P(A ∩ B′) = 0.8031 + (1 – 0.0903) – 0.7239 = 0.9889
c) P(A′ ∪ B′) = 1 – P(A ∩ B) = 1 – 0.0792 = 0.9208

2-95. Consider the bar code 39 is a common bar code system that consists of narrow and wide bars (black) separated by either wide or narrow spaces (white). Each character contains nine elements (five bars and four spaces). The code for a character starts and ends with a bar (either narrow or wide) and a (white) space appears between each bar. The original specification (since revised) used *exactly* two wide bars and one wide space in each character. For example, if *b* and *B* denote narrow and wide (black) bars, respectively, and *w* and *W* denote narrow and wide (white) spaces, a valid character is *bwBwBWbwb* (the number 6). Suppose that all 40 codes are equally likely (none is held back as a delimiter).
Determine the probability for each of the following:

(a) The first bar is wide or the second bar is wide.
(b) Neither the first nor the second bar is wide.
(c) The first bar is wide or the second bar is not wide.
(d) The first bar is wide or the first space is wide.

(a) Number of permutations of the bars with the first bar wide is 4 .
Number of permutations of the bars with the second bar wide is 4 .
Number of permutations of the bars with both the first & second bar wide is 1 .
Number of permutations of the bars with either the first bar wide or the last bar wide = 4 + 4 – 1 = 7.
Number of codes is multiplied this by the number of permutations for the spaces = 4.
P(first bar is wide) = 16/40 = 0.4, P(second bar is wide) = 16/40 = 0.4, P(first & second bar is wide) = 4/40 = 0.1
P(first or second bar is wide) = 4/10 + 4/10 – 1/10 = 7/10

(b) Neither the first or second bar wide implies the two wide bars occur in the last 3 positions.
Number of permutations of the bars with the wide bars in the last 3 positions is 3!/2!1! = 3
P(neither first nor second bar is wide) = 12/40 = 0.3

(c) The spaces are handled as in part (a).
P(first bar is wide) = 16/40 = 0.4
Number of permutations of the bars with the second bar narrow is 4!/(2!2!) = 6
P(second bar is narrow) = 24/40 = 0.6
Number of permutations with the first bar wide and the second bar narrow is 3!/(1!2!) = 3
P(first bar wide and the second bar narrow) = 12/40 = 0.3
P(first bar is wide or the second bar is narrow) = 0.4 + 0.6 – 0.3 = 0.7

(d) The spaces are handled as in part (a).
Number of permutations of the bars with the first bar wide is 4. Therefore, P(first bar is wide) = 16/40 = 0.4
The number of permutations of the bars = 10. Number of permutations of the spaces with the first space wide is 1.
Therefore, P(first space is wide) = 1(10)/40 = 0.25
Number codes with the first bar wide and the first space wide is 4(1) = 4
P(first bar wide & the first space wide) = 4/40 = 0.1
P(first bar is wide or the first space is wide) = 0.4 + 0.25 – 0.1 = 0.55

2-96. Consider the three patient groups. The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C," [*Gastroenterology* (1996, Vol. 111, no. 5, pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let *A* denote the event that the patient was treated with ribavirin plus interferon alfa, and let *B* denote the event that the response was complete. Determine the following probabilities:

(a) $P(A \cup B)$ (b) $P(A' \cup B)$ (c) $P(A \cup B')$

(a) P(AUB)= P(A) + P(B) - P(A∩B) = 21/60 + 22/60 – 16/60 = 9/20= 0.45
(b) P(A'UB)= P(A') + P(B) - P(A'∩B) = (19+20)/60 + 22/60 – 6/60 = 11/12 = 0.9166
(c) P(AUB')= P(A) + P(B') - P(A∩B')= 21/60 + (60-22)/60 – 5/60 = 9/10 = 0.9

2-97. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (*a*–*z*) or 26 uppercase letters (*A*–*Z*) or 10 integers (0–9). Assume all passwords are equally likely. Let *A* and *B* denote the events that consist of passwords with only letters or only integers, respectively. Determine the following probabilities:

(a) $P(A \cup B)$ (b) $P(A' \cup B)$ (c) *P* (Password contains exactly 1 or 2 integers)

(a) P(AUB) = P(A) + P(B) = $\frac{52^8}{62^8} + \frac{10^8}{62^8}$ = 0.245
(b) P(A'UB) = P(A') = $\frac{10^8}{62^8}$ = 1 – 0.2448 = 0.755
(c) P(contains exactly 1 integer)
Number of positions for the integer is 8!/(1!7!) = 8
Number of values for the integer = 10
Number of permutations of the seven letters is 52^7
Total number of permutations is 62^8
Therefore, the probability is
$\frac{8(10)(52^7)}{62^8} = 0.377$

P(contains exactly 2 integers)
Number of positions for the integers is 8!/(2!6!) = 28
Number of permutations of the two integers is 100
Number of permutations of the 6 letters is 52^6
Total number of permutations is 62^8
Therefore, the probability is
$\frac{28(100)(52^6)}{62^8} = 0.254$
Therefore, P(exactly one integer or exactly two integers) = 0.377 + 0.254 = 0.630

2-98. The article ["Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis," *Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381– 3390)] considered four treatment groups. The groups consisted of patients with different drug therapies (such as prednisone and infliximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or initial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease progression. The number of patients without progression of joint damage was 76 of 114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose that a patient is selected randomly. Let *A* denote the event that the patient is in group 1, and let *B* denote the event that there is no progression. Determine the following probabilities:

(a) $P(A \cup B)$ (b) $P(A' \cup B')$ (c) $P(A \cup B')$

P(A)= $\frac{114}{114+112+120+121} = \frac{114}{467} = 0.244$
P(B)= $\frac{76+82+104+113}{114+112+120+121} = \frac{375}{467} = 0.8029$
P(A∩B)= $\frac{76}{467} = 0.162$
(a) P(AUB) = P(A) + P(B) - P(A∩B)= 114/467 + 375/467 – 76/467 = 0.884
(b) P(A'UB') = 1 - (A∩B) = 1 – 76/467 =0.838
(c) P(AUB') = P(A) + P(B') - P(A∩B') = 114/467 + (1 – 375/467) – (114 – 76)/467 = 0.359

Section 2-4

2-99. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$ (c) $P(A \mid B)$ (d) $P(B \mid A)$

(a) P(A) = 86/100
(b) P(B) = 79/100
(c) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{70/100}{79/100} = \dfrac{70}{79}$
(d) $P(B|A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{70/100}{86/100} = \dfrac{70}{86}$

2-100. Samples of skin experiencing desquamation are analyzed for both moisture and melanin content. The results from 100 skin samples are as follows:

		Melanin Content	
		High	Low
Moisture	High	13	7
Content	Low	48	32

Let A denote the event that a sample has low melanin content, and let B denote the event that a sample has high moisture content. Determine the following probabilities:

(a) $P(A)$ (b) $P(B)$ (c) $P(A \mid B)$ (d) $P(B \mid A)$

(a) $P(A) = \dfrac{7+32}{100} = 0.39$

(b) $P(B) = \dfrac{13+7}{100} = 0.2$

(c) $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{7/100}{20/100} = 0.35$

(d) $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{7/100}{39/100} = 0.1795$

2-101. The analysis of results from a leaf transmutation experiment (turning a leaf into a petal) is summarized by type of transformation completed:

		Total Textural Transformation	
		Yes	No
Total Color	Yes	243	26
Transformation	No	13	18

(a) If a leaf completes the color transformation, what is the probability that it will complete the textural transformation?
(b) If a leaf does not complete the textural transformation, what is the probability it will complete the color transformation?

Let *A* denote the event that a leaf completes the color transformation and let *B* denote the event that a leaf completes the textural transformation. The total number of experiments is 300.

(a) $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{243/300}{(243+26)/300} = 0.903$

(b) $P(A \mid B') = \dfrac{P(A \cap B')}{P(B')} = \dfrac{26/300}{(18+26)/300} = 0.591$

2-102. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and length measurements. The results of 100 parts are summarized as follows:

		Length	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

Let *A* denote the event that a sample has excellent surface finish, and let *B* denote the event that a sample has excellent length. Determine:

(a) $P(A)$ (b) $P(B)$ (c) $P(A \mid B)$ (d) $P(B \mid A)$

(e) If the selected part has excellent surface finish, what is the probability that the length is excellent?

(f) If the selected part has good length, what is the probability that the surface finish is excellent?

(a) 0.82

(b) 0.90

(c) 8/9 = 0.889

(d) 80/82 = 0.9756

(e) 80/82 = 0.9756

(f) 2/10 = 0.20

2-103. The following table summarizes the analysis of samples of galvanized steel for coating weight and surface roughness:

		Coating Weight	
		High	Low
Surface	High	12	16
Roughness	Low	88	34

(a) If the coating weight of a sample is high, what is the probability that the surface roughness is high?

(b) If the surface roughness of a sample is high, what is the probability that the coating weight is high?

(c) If the surface roughness of a sample is low, what is the probability that the coating weight is low?

(a) 12/100 (b) 12/28 (c) 34/122

2-104. Consider the data on wafer contamination and location in the sputtering tool shown in Table 2-2. Assume that one wafer is selected at random from this set. Let *A* denote the event that a wafer contains four or more particles, and let *B* denote the event that a wafer is from the center of the sputtering tool. Determine:

(a) $P(A)$ (b) $P(A \mid B)$ (c) $P(B)$ (d) $P(B \mid A)$ (e) $P(A \cap B)$ (f) $P(A \cup B)$

(a) P(A) = 0.05 + 0.10 = 0.15

(b) P(A|B) = $\dfrac{P(A \cap B)}{P(B)} = \dfrac{0.04+0.07}{0.72} = 0.153$

(c) P(B) = 0.72

(d) P(B|A) = $\dfrac{P(A \cap B)}{P(A)} = \dfrac{0.04+0.07}{0.15} = 0.733$

(e) $P(A \cap B) = 0.04 + 0.07 = 0.11$

(f) $P(A \cup B) = 0.15 + 0.72 - 0.11 = 0.76$

2-105. The following table summarizes the number of deceased beetles under autolysis (the destruction of a cell after its death by the action of its own enzymes) and putrefaction (decomposition of organic matter, especially protein, by microorganisms, resulting in production of foul-smelling matter):

		Autolysis	
		High	Low
Putrefaction	High	14	59
	Low	18	9

(a) If the autolysis of a sample is high, what is the probability that the putrefaction is low?
(b) If the putrefaction of a sample is high, what is the probability that the autolysis is high?
(c) If the putrefaction of a sample is low, what is the probability that the autolysis is low?

Let *A* denote the event that autolysis is high and let *B* denote the event that putrefaction is high. The total number of experiments is 100.

(a) $P(B'|A)=\dfrac{P(A\cap B')}{P(A)}=\dfrac{18/100}{(14+18)/100}=0.5625$

(b) $P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{14/100}{(14+59)/100}=0.1918$

(c) $P(A'|B')=\dfrac{P(A'\cap B')}{P(B')}=\dfrac{9/100}{(18+9)/100}=0.333$

2-106. A maintenance firm has gathered the following information regarding the failure mechanisms for air conditioning systems:

		Evidence of Gas Leaks	
		Yes	No
Evidence of electrical failure	Yes	55	17
	No	32	3

The units without evidence of gas leaks or electrical failure showed other types of failure. If this is a representative sample of AC failure, find the probability
(a) That failure involves a gas leak
(b) That there is evidence of electrical failure given that there was a gas leak
(c) That there is evidence of a gas leak given that there is evidence of electrical failure

(a) P(gas leak) = (55 + 32)/107 = 0.813
(b) P(electric failure | gas leak) = (55/107)/(87/102) = 0.632
(c) P(gas leak | electric failure) = (55/107)/(72/107) = 0.764

2-107. A lot of 100 semiconductor chips contains 20 that are defective. Two are selected randomly, without replacement, from the lot.
(a) What is the probability that the first one selected is defective?
(b) What is the probability that the second one selected is defective given that the first one was defective?
(c) What is the probability that both are defective?
(d) How does the answer to part (b) change if chips selected were replaced prior to the next selection?

(a) 20/100
(b) 19/99
(c) (20/100)(19/99) = 0.038
(d) If the chips were replaced, the probability would be (20/100) = 0.2

2-108. A batch of 500 containers for frozen orange juice contains 5 that are defective. Two are selected, at random, without replacement from the batch.
(a) What is the probability that the second one selected is defective given that the first one was defective?
(b) What is the probability that both are defective?
(c) What is the probability that both are acceptable?
Three containers are selected, at random, without replacement, from the batch.
(d) What is the probability that the third one selected is defective given that the first and second ones selected were defective?

(e) What is the probability that the third one selected is defective given that the first one selected was defective and the second one selected was okay?
(f) What is the probability that all three are defective?

(a) 4/499 = 0.0080
(b) (5/500)(4/499) = 0.000080
(c) (495/500)(494/499) = 0.98
(d) 3/498 = 0.0060
(e) 4/498 = 0.0080
(f) $\left(\frac{5}{500}\right)\left(\frac{4}{499}\right)\left(\frac{3}{498}\right) = 4.82\text{x}10^{-7}$

2-109 A batch of 350 samples of rejuvenated mitochondria contains 8 that are mutated (or defective). Two are selected from the batch, at random, without replacement.
(a) What is the probability that the second one selected is defective given that the first one was defective?
(b) What is the probability that both are defective?
(c) What is the probability that both are acceptable?

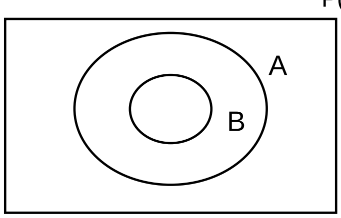
(a) P = (8-1)/(350-1)=0.020
(b) P = (8/350) $\times$ [(8-1)/(350-1)]=0.000458
(c) P = (342/350) $\times$ [(342-1)/(350-1)]=0.9547

2-110. A computer system uses passwords that are exactly seven characters and each character is one of the 26 letters (*a*–*z*) or 10 integers (0–9). You maintain a password for this computer system. Let *A* denote the subset of passwords that begin with a vowel (either *a*, *e*, *i*, *o*, or *u*) and let *B* denote the subset of passwords that end with an even number (either 0, 2, 4, 6, or 8).
(a) Suppose a hacker selects a password at random. What is the probability that your password is selected?
(b) Suppose a hacker knows that your password is in event *A* and selects a password at random from this subset. What is the probability that your password is selected?
(c) Suppose a hacker knows that your password is in *A* and *B*and selects a password at random from this subset. What is the probability that your password is selected?

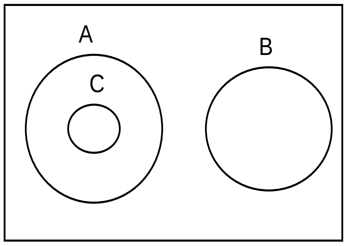
(a) $\frac{1}{36^7}$
(b) $\frac{1}{5(36^6)}$
(c) $\frac{1}{5(36^5)5}$

2-111. If $P(A \mid B) = 1$, must $A = B$? Draw a Venn diagram to explain your answer.

No, if $B \subset A$, then P(A/B) = $\frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$



2-112. Suppose *A* and *B* are mutually exclusive events. Construct a Venn diagram that contains the three events *A*, *B*, and *C* such that $P(A \mid C) = 1$ and $P(B \mid C) = 0$.



2-113. Consider the endothermic reactions given below. Let A denote the event that a reaction's final temperature is 271 K or less. Let B denote the event that the heat absorbed is above target.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Determine the following probabilities.

(a) $P(A \mid B)$ (b) $P(A' \mid B)$ (c) $P(A \mid B')$ (d) $P(B \mid A)$

(a) $$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{(40+16)/204}{(40+16+36)/204} = \frac{56}{92} = 0.6087$$

(b) $$P(A' \mid B) = \frac{P(A' \cap B)}{P(B)} = \frac{36/204}{(40+16+36)/204} = \frac{36}{92} = 0.3913$$

(c) $$P(A \mid B') = \frac{P(A \cap B')}{P(B')} = \frac{56/204}{(12+44+56)/204} = \frac{56}{112} = 0.5$$

(d) $$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{(40+16)/204}{(40+12+16+44)/204} = \frac{40+16}{112} = 0.5$$

2-114. Consider the hospital emergency room data given below. Let A denote the event that a visit is to hospital 4, and let B denote the event that a visit results in LWBS (at any hospital).

	Hospital				
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Determine the following probabilities.

(a) $P(A \mid B)$ (b) $P(A' \mid B)$ (c) $P(A \mid B')$ (d) $P(B \mid A)$

(a) $$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{242/22252}{953/22252} = \frac{242}{953} = 0.2539$$

(b) $$P(A' \mid B) = \frac{P(A' \cap B)}{P(B)} = \frac{(195+270+246)/22252}{953/22252} = \frac{711}{953} = 0.7461$$

(c) $P(A|B') = \dfrac{P(A \cap B')}{P(B')} = \dfrac{(984+3103)/22252}{(22252-953)/22252} = \dfrac{4087}{21299} = 0.1919$

(d) $P(B|A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{242/22252}{4329/22252} = \dfrac{242}{4329} = 0.0559$

2-115. Consider the well failure data given below.

	Wells	
Geological Formation Group	**Failed**	**Total**
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

(a) What is the probability of a failure given there are more than 1,000 wells in a geological formation?
(b) What is the probability of a failure given there are fewer than 500 wells in a geological formation?

(a) $P(B|A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{(170+443+60)/8493}{(1685+3733+1403)/8493} = \dfrac{673}{6821} = 0.0987$

Also the probability of failure for fewer than 1000 wells is

$$P(B|A') = \frac{P(B \cap A')}{P(B')} = \frac{(2+14+29+46+3)/8493}{(28+363+309+933+39)/8493} = \frac{92}{1672} = 0.0562$$

(b) Let C denote the event that fewer than 500 wells are present.

$$P(B|C) = \frac{P(A \cap C)}{P(C)} = \frac{(2+14+29+46+3)/8493}{(28+363+309+39)/8493} = \frac{48}{739} = 0.0650$$

2-116. An article in the *The Canadian Entomologist* (Harcourt et al., 1977, Vol. 109, pp. 1521–1534) reported on the life of the alfalfa weevil from eggs to adulthood. The following table shows the number of larvae that survived at each stage of development from eggs to adults.

Eggs	Early Larvae	Late Larvae	Pre-pupae	Late Pupae	Adults
421	412	306	45	35	31

(a) What is the probability an egg survives to adulthood?
(b) What is the probability of survival to adulthood given survival to the late larvae stage?
(c) What stage has the lowest probability of survival to the next stage?

Let A denote the event that an egg survives to an adult
Let EL denote the event that an egg survives at early larvae stage
Let LL denote the event that an egg survives at late larvae stage
Let PP denote the event that an egg survives at pre-pupae larvae stage
Let LP denote the event that an egg survives at late pupae stage

(a) $P(A) = 31/421 = 0.0736$

(b) $P(A|LL) = \dfrac{P(A \cap LL)}{P(LL)} = \dfrac{31/421}{306/421} = 0.1013$

(c) $P(EL) = 412/421 = 0.9786$

$$P(LL|EL) = \frac{P(LL \cap EL)}{P(EL)} = \frac{306/421}{412/421} = 0.7427$$

$$P(PP \mid LL) = \frac{P(PP \cap LL)}{P(LL)} = \frac{45/421}{306/421} = 0.1471$$

$$P(LP \mid PP) = \frac{P(LP \cap PP)}{P(PP)} = \frac{35/421}{45/421} = 0.7778$$

$$P(A \mid LP) = \frac{P(A \cap LP)}{P(LP)} = \frac{31/421}{35/421} = 0.8857$$

The late larvae stage has the lowest probability of survival to the pre-pupae stage.

2-117. Consider the bar code 39 is a common bar code system that consists of narrow and wide bars (black) separated by either wide or narrow spaces (white). Each character contains nine elements (five bars and four spaces). The code for a character starts and ends with a bar (either narrow or wide) and a (white) space appears between each bar. The original specification (since revised) used *exactly* two wide bars and one wide space in each character. For example, if *b* and *B* denote narrow and wide (black) bars, respectively, and *w* and *W* denote narrow and wide (white) spaces, a valid character is *bwBwBWbwb* (the number 6).. Suppose that all 40 codes are equally likely (none is held back as a delimiter).
Determine the probability for each of the following:

(a) The second bar is wide given that the first bar is wide.
(b) The third bar is wide given that the first two bars are not wide.
(c) The first bar is wide given that the last bar is wide.

(a) A = permutations with first bar wide, B = permutations with second bar wide

P(B|A) = $\frac{P(A \cap B)}{P(A)}$

There are 5 bars and 2 are wide. Number of permutations of the bars with 2 wide and 3 narrow bars is 5!/(2!3!) = 10
Number of permutations of the 4 spaces is 4!/(1!3!) = 4

Number of permutations of the bars with the first bar wide is 4!/(3!1!) = 4. Spaces are handled as previously. Therefore, P(A) = 16/40 = 0.4
Number of permutations of the bars with the first and second bar wide is 1. Spaces are handled as previously. Therefore, P(A ∩ B) = 4/40 = 0.1
Therefore, P(B|A) = 0.1/0.4 = 0.25
Or can use the fact that if the first bar is wide there are 4 other equally likely positions for the wide bar. Therefore, P(B|A) = 0.25

(b) A = first two bars not wide, B = third bar wide

P(B|A) = $\frac{P(A \cap B)}{P(A)}$

Number of permutations of the bars with the first two bars not wide is 3!/2!1! = 3. Spaces are handled as in part (a). Therefore, P(A) = 12/40 =0.3
Number of permutations of the bars with the first two bars not wide and the third bar wide is 2. Spaces are handled as in part (a). Therefore, P(A ∩ B) = 8/40
Therefore, P(B|A) = 0.2/0.3 = 2/3
Or can use the fact that if the first two bars are not wide, there are only 3 equally likely positions for the 2 wide bars and 2 of these positions result in a wide bar in the third position. Therefore, P(B|A) = 2/3

(c) A = first bar wide, B = last bar wide

P(A|B) = $\frac{P(A \cap B)}{P(B)}$

Number of permutations of the bars with last bar wide is 4!/(3!1!) = 4. Spaces are handled as in part (a). Therefore, P(B) = 16/40 = 0.4
Number of permutations of the bar with the first and last bar wide is 1. Spaces are handled as in part (a). Therefore, P(A ∩ B) = 4/40 = 0.1 and P(A|B) = 0.1/0.4 = 0.25

Or can use the fact that if the last bar is wide there are 4 other equally likely positions for the wide bar. Therefore, P(B|A) = 0.25

2-118. The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C," [*Gastroenterology* (1996, Vol. 111, no. 5, pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let *A* denote the event that the patient is treated with ribavirin plus interferon alfa, and let *B* denote the event that the response is complete. Determine the following probabilities:

(a) $P(B \mid A)$ (b) $P(A \mid B)$ (c) $P(A \mid B')$ (d) $P(A' \mid B)$

P(A) = 21/60= 0.35, P(B) = 22/60 = 0.366, P(A∩B) = 16/60 = 0.266

(a) P(B|A)= $\frac{P(A\cap B)}{P(A)} = \frac{16/60}{21/60} = 0.762$

(b) P(A|B)= $\frac{P(A\cap B)}{P(B)} = \frac{16/60}{22/60} = 0.727$

(c) P(A|B')= $\frac{P(A\cap B')}{P(B')} = \frac{5/60}{38/60} = 0.131$

(d) P(A'|B)= $\frac{P(A'\cap B)}{P(B)} = \frac{6/60}{22/60} = 0.272$

2-119. The article ["Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis," *Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381– 3390)] considered four treatment groups. The groups consisted of patients with different drug therapies (such as prednisone and infliximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or initial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease progression. The number of patients without progression of joint damage was 76 of 114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose that a patient is selected randomly. Let *A* denote the event that the patient is in group 1, and let *B* denote the event that there is no progression. Determine the following probabilities:

(a) $P(B \mid A)$ (b) $P(A \mid B)$ (c) $P(A \mid B')$ (d) $P(A' \mid B)$

P(A) = 114/467 P(B) = 375/467 P(A ∩ B) = 76/467

(a) P(B|A)= $\frac{P(A\cap B)}{P(A)} = \frac{76/467}{114/467} = 0.667$

(b) P(A|B)= $\frac{P(A\cap B)}{P(B)} = \frac{76/467}{375/467} = 0.203$

(c) P(A|B')= $\frac{P(A\cap B')}{P(B')} = \frac{38/467}{92/467} = 0.413$

(d) P(A'|B)= $\frac{P(A'\cap B)}{P(B)} = \frac{299/467}{375/467} = 0.797$

2-120. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (*a*–*z*) or 26 uppercase letters (*A*–*Z*) or 10 integers (0–9). Let Ω denote the set of all possible passwords. Suppose that all passwords in Ω are equally likely. Determine the probability for each of the following:
(a) Password contains all lowercase letters given that it contains only letters
(b) Password contains at least 1 uppercase letter given that it contains only letters
(c) Password contains only even numbers given that is contains all numbers

Let A= passwords with all letters, B = passwords with all lowercase letters

(a) P(B|A)= $\frac{P(A\cap B)}{P(A)} = \frac{\frac{26^8}{62^8}}{\frac{52^8}{62^8}} = \frac{26^8}{52^8} = 0.0039$

(b) C = passwords with at least 1 uppercase letter

P(C|A)= $\frac{P(A\cap C)}{P(A)}$

P(A ∩ C) = P(A) – P(A ∩ C') = $\frac{52^8}{62^8} - \frac{26^8}{62^8}$

P(A) = $\frac{52^8}{62^8}$

Therefore, P(C | A) = 1 - $\frac{26^8}{52^8}$=0.996

(c) P(containing all even numbers | contains all numbers) = $\frac{5^8}{10^8} = 0.0039$

Section 2-5

2-121. Suppose that $P(A \mid B) = 0.4$ and $P(B) = 0.5$. Determine the following:

(a) $P(A \cap B)$ (b) $P(A' \cap B)$

(a) $P(A \cap B) = P(A|B)P(B) = (0.4)(0.5) = 0.20$

(b) $P(A' \cap B) = P(A'|B)P(B) = (0.6)(0.5) = 0.30$

2-122. Suppose that $P(A \mid B) = 0.2$, $P(A \mid B') = 0.3$, and $P(B) = 0.8$. What is $P(A)$?

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B') \\ &= P(A|B)P(B) + P(A|B')P(B') \\ &= (0.2)(0.8) + (0.3)(0.2) \\ &= 0.16 + 0.06 = 0.22 \end{aligned}$$

2-123. The probability is 1% that an electrical connector that is kept dry fails during the warranty period of a portable computer. If the connector is ever wet, the probability of a failure during the warranty period is 5%. If 90% of the connectors are kept dry and 10% are wet, what proportion of connectors fail during the warranty period? Let F denote the event that a connector fails and let W denote the event that a connector is wet.

$$\begin{aligned} P(F) &= P(F|W)P(W) + P(F|W')P(W') \\ &= (0.05)(0.10) + (0.01)(0.90) = 0.014 \end{aligned}$$

2-124. Suppose 2% of cotton fabric rolls and 3% of nylon fabric rolls contain flaws. Of the rolls used by a manufacturer, 70% are cotton and 30% are nylon. What is the probability that a randomly selected roll used by the manufacturer contains flaws?

Let F denote the event that a roll contains a flaw and let C denote the event that a roll is cotton.

$$\begin{aligned} P(F) &= P(F|C)P(C) + P(F|C')P(C') \\ &= (0.02)(0.70) + (0.03)(0.30) = 0.023 \end{aligned}$$

2-125. The edge roughness of slit paper products increases as knife blades wear. Only 1% of products slit with new blades have rough edges, 3% of products slit with blades of average sharpness exhibit roughness, and 5% of products slit with worn blades exhibit roughness. If 25% of the blades in manufacturing are new, 60% are of average sharpness, and 15% are worn, what is the proportion of products that exhibit edge roughness?

Let R denote the event that a product exhibits surface roughness. Let N, A, and W denote the events that the blades are new, average, and worn, respectively. Then,

P(R)= P(R|N)P(N) + P(R|A)P(A) + P(R|W)P(W)
= (0.01)(0.25) + (0.03) (0.60) + (0.05)(0.15)
= 0.028

2-126. In the 2012 presidential election, exit polls from the critical state of Ohio provided the following results:

Total	Obama	Romney
No college degree (60%)	52%	45%
College graduate (40%)	47%	51%

What is the probability a randomly selected respondent voted for Obama?

Let A denote the event that a respondent is a college graduate and let B denote the event that an individual votes for Obama.
$P(B) = P(A)P(B|A) + P(A')P(B|A') = 0.40\times 0.47+0.60\times 0.52 = 0.50$

2-127. Computer keyboard failures are due to faulty electrical connects (12%) or mechanical defects (88%). Mechanical defects are related to loose keys (27%) or improper assembly (73%). Electrical connect defects are caused by defective wires (35%), improper connections (13%), or poorly welded wires (52%).
(a) Find the probability that a failure is due to loose keys.
(b) Find the probability that a failure is due to improperly connected or poorly welded wires.

(a) (0.88)(0.27) = 0.2376
(b) (0.12)(0.13+0.52) = 0.0.078

2-128. Heart failures are due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances (73%) or foreign objects (27%). Natural occurrences are caused by arterial blockage (56%), disease (27%), and infection (e.g., staph infection) (17%).
(a) Determine the probability that a failure is due to an induced substance.
(b) Determine the probability that a failure is due to disease or infection

(a)$P = 0.13\times 0.73=0.0949$
(b)$P = 0.87\times (0.27+0.17)=0.3828$

2-129. A batch of 25 injection-molded parts contains 5 parts that have suffered excessive shrinkage.
(a) If two parts are selected at random, and without replacement, what is the probability that the second part selected is one with excessive shrinkage?
(b) If three parts are selected at random, and without replacement, what is the probability that the third part selected is one with excessive shrinkage?

Let A and B denote the event that the first and second part selected has excessive shrinkage, respectively.
(a) $P(B)= P(B|A)P(A) + P(B|A')P(A')$

$= (4/24)(5/25) + (5/24)(20/25) = 0.20$
(b) Let C denote the event that the third part selected has excessive shrinkage.

$$P(C) = P(C|A\cap B)P(A\cap B) + P(C|A\cap B')P(A\cap B')$$
$$+P(C|A'\cap B)P(A'\cap B) + P(C|A'\cap B')P(A'\cap B')$$
$$= \frac{3}{23}\left(\frac{4}{24}\right)\left(\frac{5}{25}\right)+\frac{4}{23}\left(\frac{20}{24}\right)\left(\frac{5}{25}\right)+\frac{4}{23}\left(\frac{5}{24}\right)\left(\frac{20}{25}\right)+\frac{5}{23}\left(\frac{19}{24}\right)\left(\frac{20}{25}\right)$$
$$= 0.20$$

2-130. A lot of 100 semiconductor chips contains 20 that are defective.
(a) Two are selected, at random, without replacement, from the lot. Determine the probability that the second chip selected is defective.
(b) Three are selected, at random, without replacement, from the lot. Determine the probability that all are defective.

Let A and B denote the events that the first and second chips selected are defective, respectively.
(a) $P(B) = P(B|A)P(A) + P(B|A')P(A') = (19/99)(20/100) + (20/99)(80/100) = 0.2$
(b) Let C denote the event that the third chip selected is defective.

$$P(A \cap B \cap C) = P(C|A \cap B)P(A \cap B) = P(C|A \cap B)P(B|A)P(A)$$
$$= \frac{18}{98}\left(\frac{19}{99}\right)\left(\frac{20}{100}\right)$$
$$= 0.00705$$

2-131. An article in the *British Medical Journal* ["Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy" (1986, Vol. 82, pp. 879–892)] provided the following discussion of success rates in kidney stone removals. Open surgery had a success rate of 78% (273/350) and a newer method, percutaneous nephrolithotomy (PN), had a success rate of 83% (289/350). This newer method looked better, but the results changed when stone diameter was considered. For stones with diameters less than 2 centimeters, 93% (81/87) of cases of open surgery were successful compared with only 83% (234/270) of cases of PN. For stones greater than or equal to 2 centimeters, the success rates were 73% (192/263) and 69% (55/80) for open surgery and PN, respectively. Open surgery is better for both stone sizes, but less successful in total. In 1951, E. H. Simpson pointed out this apparent contradiction (known as **Simpson's paradox**), and the hazard still persists today. Explain how open surgery can be better for both stone sizes but worse in total.

Open surgery

	success	failure	sample size	sample percentage	conditional success rate
large stone	192	71	263	75%	73%
small stone	81	6	87	25%	93%
overall summary	273	77	350	100%	78%

PN

	success	failure	sample size	sample percentage	conditional success rate
large stone	55	25	80	23%	69%
small stone	234	36	270	77%	83%
overall summary	289	61	350	100%	83%

The overall success rate depends on the success rates for each stone size group, but also the probability of the groups. It is the weighted average of the group success rate weighted by the group size as follows

P(overall success) = P(success| large stone)P(large stone)) + P(success| small stone)P(small stone).

For open surgery, the dominant group (large stone) has a smaller success rate while for PN, the dominant group (small stone) has a larger success rate.

2-132. Consider the endothermic reactions given below. Let *A* denote the event that a reaction's final temperature is 271 K or less. Let *B* denote the event that the heat absorbed is above target.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Determine the following probabilities.

(a) $P(A \cap B)$ (b) $P(A \cup B)$ (c) $P(A' \cup B')$ (d) Use the total probability rule to determine $P(A)$

P(A) = 112/204 = 0.5490, P(B) = 92/204 = 0.4510
(a) P(A ∩ B) = P(A | B)P(B) = (56/92)(92/204) = 0.2745
(b) P(A ∪ B) = P(A) + P(B) – P(A ∩ B) = 0.5490 + 0.4510 – 0.2745 = 0.7255
(c) P(A′ ∪ B′) = 1 – P(A ∩ B) = 1 – 0.2745 = 0.7255
(d) P(A) = P(A | B)P(B) + P(A | B')P(B') = (56/92) (92/204) + (56/112)(112/204) = 112/204 = 0.5490

2-133. Consider the hospital emergency room data given below. Let *A* denote the event that a visit is to hospital 4 and let *B* denote the event that a visit results in LWBS (at any hospital).

Hospital					
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Determine the following probabilities.

(a) $P(A \cap B)$ (b) $P(A \cup B)$ (c) $P(A' \cup B')$ (d)Use the total probability rule to determine $P(A)$

P(A) = 4329/22252 = 0.1945, P(B) = 953/22252 = 0.0428
(a) $P(A \cap B) = P(A \mid B)P(B) = (242/953)(953/22252) = 0.0109$
(b) $P(A \cup B) = P(A) + P(B) – P(A \cap B) = 0.1945 + 0.0428 – 0.0109 = 0.2264$
(c) $P(A' \cup B') = 1 – P(A \cap B) = 1 – 0.0109 = 0.9891$
(d) P(A) = P(A | B)P(B) + P(A | B')P(B') = (242/953)(953/22252) + (4087/21299)(21299/22252) = 0.1945

2-134. Consider the hospital emergency room data given below. Suppose that three visits that resulted in LWBS are selected randomly (without replacement) for a follow-up interview.

Hospital					
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

(a) What is the probability that all three are selected from hospital 2?
(b) What is the probability that all three are from the same hospital?

(a) $P = \dfrac{\binom{270}{3}}{\binom{953}{3}} = 0.0226$

(b) $P = \dfrac{\binom{195}{3} + \binom{270}{3} + \binom{246}{3} + \binom{242}{3}}{\binom{953}{3}} = 0.0643$

2-135. Consider the well failure data given below. Let *A* denote the event that the geological formation has more than 1000 wells, and let *B* denote the event that a well failed.

	Wells	
Geological Formation Group	Failed	Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

Determine the following probabilities.

(a) $P(A\_ B)$ (b) $P(A\_ B)$ (c) $P(A\_\_ B\_)$ (d) Use the total probability rule to determine $P(A)$

P(A) = (1685 + 3733 + 1403)/8493 = 0.8031, P(B) = (170 + 2 + 443 + 14 + 29 + 60 + 46 + 3)/8493 = 0.0903

(a) $P(A \cap B) = P(B \mid A)P(A) = (673/6821)(6821/8493) = 0.0792$
(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8031 + 0.0903 - 0.0792 = 0.8142$
(c) $P(A' \cup B') = 1 - P(A \cap B) = 1 - 0.0792 = 0.9208$
(d) $P(A) = P(A \mid B)P(B) + P(A \mid B')P(B') = (673/767)(767/8493) + (6148/7726)(7726/8493) = 0.8031$

2-136. Consider the well failure data given below. Suppose that two failed wells are selected randomly (without replacement) for a follow-up review.

	Wells	
Geological Formation Group	Failed	Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

(a) What is the probability that both are from the gneiss geological formation group?
(b) What is the probability that both are from the same geological formation group?

(a) $$P = \frac{\binom{170}{2}}{\binom{767}{2}} = 0.0489$$

(b) $$P = \frac{\binom{170}{2}+\binom{2}{2}+\binom{443}{2}+\binom{14}{2}+\binom{29}{2}+\binom{60}{2}+\binom{46}{2}+\binom{3}{2}}{\binom{767}{2}} = 0.3934$$

2-137. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Determine the probability that the ad color is red and the font size is not the smallest one.

Let R denote red color and F denote that the font size is not the smallest. Then $P(R) = 1/4$, $P(F) = 4/5$.
Because the Web sites are generated randomly these events are independent. Therefore, $P(R \cap F) = P(R)P(F) = (1/4)(4/5) = 0.2$

2-138. Consider the code 39 is a common bar code system that consists of narrow and wide bars (black) separated by either wide or narrow spaces (white). Each character contains nine elements (five bars and four spaces). The code for a character starts and ends with a bar (either narrow or wide) and a (white) space appears between each bar. The original specification (since revised) used *exactly* two wide bars and one wide space in each character. For example, if *b* and *B* denote narrow and wide (black) bars, respectively, and *w* and *W* denote narrow and wide (white) spaces, a valid character is *bwBwBWbwb* (the number 6). Suppose that all 40 codes are equally likely (none is held back as a delimiter).
Determine the probability for each of the following:

(a) The code starts and ends with a wide bar.
(b) Two wide bars occur consecutively.
(c) Two consecutive wide bars occur at the start or end.
(d) The middle bar is wide.

(a) Number of permutations of the bars that start and end with a wide bar is 1. Number of permutations of the spaces is $4!/(1!3!) = 4$.
Number of codes that start and end with a wide bar = 4. P(code starts and ends with a wide bar) = 4/40 = 0.1
(b) Number of permutations of the bars where two wide bars are consecutive = 4. Spaces are handled as in part (a).
P(two wide bars are consecutive) =16/40 = 0.4

(c) Number of permutations of the bars with two consecutive wide bars at the start or end = 2. Spaces are handled as in part (a). P(two consecutive wide bars at the start or end) = 8/40 = 0.2
(d) Number of permutations of the bars with the middle bar wide = 4. Spaces are handled as in part (a). P(middle bar is wide) = 16/40 = 0.4

2-139. A hospital operating room needs to schedule three knee surgeries and two hip surgeries in a day. Suppose that an operating room needs to schedule three knee, four hip, and five shoulder surgeries. Assume that all schedules are equally likely. Determine the following probabilities:
(a) All hip surgeries are completed first given that all knee surgeries are last.
(b) The schedule begins with a hip surgery given that all knee surgeries are last.
(c) The first and last surgeries are hip surgeries given that knee surgeries are scheduled in time periods 2 through 4.
(d) The first two surgeries are hip surgeries given that all knee surgeries are last.

(a) P(all hip surgeries are completed first given that all knee surgeries are last)
A = schedules with all hip surgeries completed first
B = schedules with all knee surgeries last
Total number of schedules $=\frac{12!}{3!4!5!}$
Number of schedules with all knee surgeries last = $\frac{9!}{4!5!}$
Number of schedules with all hip surgeries first and all knee surgeries last = 1
P(B) = $\frac{9!}{4!5!}/\frac{12!}{3!4!5!}$, P(A ∩ B) = $1/\frac{12!}{3!4!5!}$
Therefore, P(A | B) = P(A ∩ B)/P(B) = $1/\frac{9!}{4!5!}$=1/126
Alternatively, one can reason that when all knee surgeries are last, there are $\frac{9!}{4!5!}$ remaining schedules and one of these has all knee surgeries first. Therefore, the solution is $1/\frac{9!}{4!5!} = 1/126$

(b) P(schedule begins with a hip surgery given that all knee surgeries are last)
C = schedules that begin with a hip surgery
B = schedules with all knee surgeries last
P(B) = $\frac{9!}{4!5!}/\frac{12!}{3!4!5!}$, P(C ∩ B) = $\frac{8!}{3!5!}/\frac{12!}{3!4!5!}$
Therefore, P(C | B) = P(C ∩ B)/P(B) = $\frac{8!}{3!5!}/\frac{9!}{4!5!}$ = 4/9
Alternatively, one can reason that when all knee surgeries are last, there are 4 hip and 5 shoulder surgeries that remain to schedule. The probability the first one is a hip surgery is then 4/9

(c) P(first and last surgeries are hip surgeries given that knee surgeries are scheduled in time periods 2 through 4)
D = schedules with first and last hip surgeries
E = schedules with knee surgeries in periods 2 through 4
P(E) $=\frac{9!}{4!5!}/\frac{12!}{3!4!5!}$
P(D∩E) = $\frac{7!}{2!5!}/\frac{12!}{3!4!5!}$
P(D | E) = P(D∩E)/P(E) = $\frac{7!}{2!5!}/\frac{9!}{4!5!}$ = 1/6
Alternatively, one can conclude that with knee surgeries in periods 2 through 4, there are $\frac{9!}{4!5!}$ remaining schedules and $\frac{7!}{2!5!}$ of these have hip surgeries first and last. Therefore, P(D | E) = P(D∩E)/P(E) = $\frac{7!}{2!5!}/\frac{9!}{4!5!}$ = 1/6

(d) P(first two surgeries are hip surgeries given that all knee surgeries are last)
F = schedules with the first two surgeries as hip surgeries
B = schedules with all knee surgeries last
P(B) = $\frac{9!}{4!5!}/\frac{12!}{3!4!5!}$,
P(F∩B) = $\frac{7!}{2!5!}/\frac{12!}{3!4!5!}$
P(F | B) = P(F∩B)/P(B) = $\frac{7!}{2!5!}/\frac{9!}{4!5!}$ = 1/6
Alternatively, one can conclude that with knee surgeries last, there are $\frac{9!}{4!5!}$ remaining schedules and $\frac{7!}{2!5!}$ have hip surgeries in the first two periods. Therefore, P(F | B) = P(F∩B)/P(B) = $\frac{7!}{2!5!}/\frac{9!}{4!5!}$ = 1/6

2-140. The article ["Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis," *Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381– 3390)] considered four treatment groups.

The groups consisted of patients with different drug therapies (such as prednisone and infliximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or initial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease progression. The number of patients without progression of joint damage was 76 of 114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose that a patient is selected randomly. Let *A* denote the event that the patient is in group 1, and let *B* denote the event for which there is no progression. Determine the following probabilities:

(a) *P*(*A*) *B*) (b) *P*(*B*) (c) *P*(*A*__ *B*) (d) *P*(*A*∗ *B*) (e) *P*(*A*__ *B*)

A = group 1, B = no progression
(a) P(A∩B) = P(B|A)P(A)= (76/114)(114/467) = 0.162
(b) P(B) = P(B|G1)P(G1) + P(B|G2)P(G2) + P(B|G3)P(G3) + P(P|G4)P(G4) = 0.802
(c) P(A'∩B) = P(B|A') P(A') = (299/353)(353/467) = 0.6403
(d) P(A U B) = P(A) + P(B) - P(A∩B) = 114/467 + 375/467 – 76/467 = 0.884
(e) P(A' U B) = P(A') + P(B) - P(A'∩B) = 353/467+375/467 – 299/467 = 0.919

2-141. A computer system uses passwords that contain exactly eight characters, and each character is one of the 26 lowercase letters (*a*–*z*) or 26 uppercase letters (*A*–*Z*) or 10 integers (0–9). Let Ω denote the set of all possible password, and let *A* and *B* denote the events that consist of passwords with only letters or only integers, respectively. Suppose that all passwords in Ω are equally likely. Determine the following probabilities:

(a) *P*(*A*|*B*2)
(b) *P*(*A*__ *B*)
(c) *P* (password contains exactly 2 integers given that it contains at least 1 integer)

A = all letters, B = all integers
(a) P(A|B') = P(A∩B')/P(B') = P(A)/P(B') = P(A)/[1 - P(B)]
P(A)= $\frac{52^8}{62^8}$, P(B)= $\frac{10^8}{62^8}$
P(A|B') = $\frac{\frac{52^8}{62^8}}{1-\frac{10^8}{62^8}}$ = 0.245
(b) P(A' ∩ B') = P(A'|B')P(B')
From part (a), P(B') = 1 – P(B) = 1 - $\frac{10^8}{62^8}$ and P(A'|B') = 1 - P(A|B') = $1-\frac{\frac{52^8}{62^8}}{1-\frac{10^8}{62^8}}$
Therefore P(A' ∩ B') = $1-\frac{52^8}{62^8}-\frac{10^8}{62^8}$ = 0.755
This can also be solved as P(A' ∩ B') = 1 - P(A U B) = $1-\frac{52^8}{62^8}-\frac{10^8}{62^8}$ because A and B are mutually exclusive.

(c) Let C = passwords with exactly 2 integers
Let D = passwords with at least one integer
P(C|D) = P(C ∩ D) / P(D) = P(C) / P(D)

P(C):
Number of positions for the integers is 8!/(2!6!) = 28
Number of ordering of the two integers is 10^2 = 100
Number of orderings of the six letters is 52^6
Total number of orderings is 62^8
Therefore, the probability is
$\frac{28(100)(52^6)}{62^8} = 0.254$
P(D):
1 – P(D) = $(52/62)^8$
P(D) = 1 - $(52/62)^8$ = 0.755

P(C | D) = 0.254/0.755 = 0.336

Section 2-6

2-142. If $P(A \mid B) = 0.4$, $P(B) = 0.8$, and $P(A) = 0.5$, are the events A and B independent?

Because P(A | B) $\neq$ P(A), the events are not independent.

2-143. If $P(A \mid B) = 0.3$, $P(B) = 0.8$, and $P(A) = 0.3$, are the events B and the complement of A independent?

P(A') = 1 – P(A) = 0.7 and P($A'|B$) = 1 – P(A | B) = 0.7

Therefore, A' and B are independent events.

2-144. If $P(A) = 0.2$, $P(B) = 0.2$, and A and B are mutually exclusive, are they independent?

If A and B are mutually exclusive, then P($A \cap B$) = 0 and P(A)P(B) = 0.04.
Therefore, A and B are <u>not</u> independent.

2-145. A batch of 500 containers of frozen orange juice contains 5 that are defective. Two are selected, at random, without replacement, from the batch. Let A and B denote the events that the first and second containers selected are defective, respectively.

(a) Are A and B independent events?
(b) If the sampling were done with replacement, would A and B be independent?

(a) $P(B \mid A) = 4/499$ and

$$P(B) = P(B|A)P(A) + P(B|A')P(A') = (4/499)(5/500) + (5/499)(495/500) = 5/500$$

Therefore, A and B are <u>not</u> independent.

(b) A and B are independent.

2-146. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:.

		Shock Resistance	
		High	Low
Scratch	High	70	9
Resistance	Low	16	5

Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance. Are events A and B independent?

P($A \cap B$) = 70/100, P(A) = 86/100, P(B) = 77/100.
Then, P($A \cap B$) $\neq$ P(A)P(B), so A and B are <u>not</u> independent.

2-147. Samples of emissions from three suppliers are classified for conformance to air-quality specifications. The results from 100 samples are summarized as follows:

		Conforms	
		Yes	No
	1	22	8
Supplier	2	25	5
	3	30	10

Let A denote the event that a sample is from supplier 1, and let B denote the event that a sample conforms to specifications.

(a) Are events A and B independent? (b) Determine $P(B \mid A)$.

(a) P($A \cap B$)= 22/100, P(A) = 30/100, P(B) = 77/100, Then P($A \cap B$) $\neq$ P(A)P(B)
Therefore, A and B are not independent.
(b) P(B|A) = P(A $\cap$ B)/P(A) = (22/100)/(30/100) = 0.733

2-148. Redundant array of inexpensive disks (RAID) is a technology that uses multiple hard drives to increase the speed of data transfer and provide instant data backup. Suppose that the probability of any hard drive failing in a day is 0.001 and the drive failures are independent.
(a) A RAID 0 scheme uses two hard drives, each containing a mirror image of the other. What is the probability of data loss? Assume that data loss occurs if both drives fail within the same day.
(b) A RAID 1 scheme splits the data over two hard drives. What is the probability of data loss? Assume that data loss occurs if at least one drive fails within the same day.

(a) $P = (0.001)^2 = 10^{-6}$

(b) $P = 1-(0.999)^2 = 0.002$

2-149. The probability that a lab specimen contains high levels of contamination is 0.10. Five samples are checked, and the samples are independent.
(a) What is the probability that none contain high levels of contamination?
(b) What is the probability that exactly one contains high levels of contamination?
(c) What is the probability that at least one contains high levels of contamination?

It is useful to work one of these exercises with care to illustrate the laws of probability. Let H_i denote the event that the *i*th sample contains high levels of contamination.

(a) $P(H_1' \cap H_2' \cap H_3' \cap H_4' \cap H_5') = P(H_1')P(H_2')P(H_3')P(H_4')P(H_5')$

by independence. Also, $P(H_i') = 0.9$. Therefore, the answer is $0.9^5 = 0.59$

(b) $A_1 = (H_1 \cap H_2' \cap H_3' \cap H_4' \cap H_5')$

$A_2 = (H_1' \cap H_2 \cap H_3' \cap H_4' \cap H_5')$

$A_3 = (H_1' \cap H_2' \cap H_3 \cap H_4' \cap H_5')$

$A_4 = (H_1' \cap H_2' \cap H_3' \cap H_4 \cap H_5')$

$A_5 = (H_1' \cap H_2' \cap H_3' \cap H_4' \cap H_5)$

The requested probability is the probability of the union $A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$ and these events are mutually exclusive. Also, by independence $P(A_i) = 0.9^4(0.1) = 0.0656$. Therefore, the answer is 5(0.0656) = 0.328.

(c) Let B denote the event that no sample contains high levels of contamination. The requested probability is P(B') = 1 - P(B). From part (a), P(B') = 1 - 0.59 = 0.41.

2-150. In a test of a printed circuit board using a random test pattern, an array of 10 bits is equally likely to be 0 or 1. Assume the bits are independent.
(a) What is the probability that all bits are 1s?
(b) What is the probability that all bits are 0s?
(c) What is the probability that exactly 5 bits are 1s and 5 bits are 0s?

Let A_i denote the event that the ith bit is a one.

(a) By independence $P(A_1 \cap A_2 \cap \ldots \cap A_{10}) = P(A_1)P(A_2)\ldots P(A_{10}) = (\frac{1}{2})^{10} = 0.000976$

(b) By independence, $P(A_1' \cap A_2' \cap \ldots \cap A_{10}') = P(A_1')P(A_2')\ldots P(A_{10}^c) = (\frac{1}{2})^{10} = 0.000976$

(c) The probability of the following sequence is

$P(A_1' \cap A_2' \cap A_3' \cap A_4' \cap A_5' \cap A_6 \cap A_7 \cap A_8 \cap A_9 \cap A_{10}) = (\frac{1}{2})^{10}$, by independence. The number of sequences consisting of five "1"'s, and five "0"'s is $\binom{10}{5} = \frac{10!}{5!5!} = 252$. The answer is $252\left(\frac{1}{2}\right)^{10} = 0.246$

2-151. Six tissues are extracted from an ivy plant infested by spider mites. The plant in infested in 20% of its area. Each tissue is chosen from a randomly selected area on the ivy plant.
(a) What is the probability that four successive samples show the signs of infestation?

(b) What is the probability that three out of four successive samples show the signs of infestation?

(a) Let I and G denote an infested and good sample. There are 3 ways to obtain four consecutive samples showing the signs of the infestation: IIIIGG, GIIIIG, GGIIII. Therefore, the probability is $3\times(0.2^4 0.8^2) = 0.003072$

(b) There are 10 ways to obtain three out of four consecutive samples showing the signs of infestation. The probability is $10\times(0.2^3 * 0.8^3) = 0.04096$

2-152. A player of a video game is confronted with a series of four opponents and an 80% probability of defeating each opponent. Assume that the results from opponents are independent (and that when the player is defeated by an opponent the game ends).

(a) What is the probability that a player defeats all four opponents in a game?
(b) What is the probability that a player defeats at least two opponents in a game?
(c) If the game is played three times, what is the probability that the player defeats all four opponents at least once?

(a) $P = (0.8)^4 = 0.4096$

(b) $P = 1 - 0.2 - 0.8\times 0.2 = 0.64$

(c) Probability defeats all four in a game = 0.8^4 = 0.4096. Probability defeats all four at least once = $1 - (1 - 0.4096)^3$ = 0.7942

2-153. In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Because acids and bases are usually colorless (as are the water and salt produced in the neutralization reaction), pH is measured to monitor the reaction. Suppose that the equivalence point is reached after approximately 100 mL of an NaOH solution has been added (enough to react with all the acetic acid present) but that replicates are equally likely to indicate from 95 to 104 mL, measured to the nearest mL. Assume that two technicians each conduct titrations independently.

(a) What is the probability that both technicians obtain equivalence at 100 mL?
(b) What is the probability that both technicians obtain equivalence between 98 and 104 mL (inclusive)?
(c) What is the probability that the average volume at equivalence from the technicians is 100 mL?

(a) The probability that one technician obtains equivalence at 100 mL is 0.1.

So the probability that both technicians obtain equivalence at 100 mL is $0.1^2 = 0.01$.

(b) The probability that one technician obtains equivalence between 98 and 104 mL is 0.7.

So the probability that both technicians obtain equivalence between 98 and 104 mL is $0.7^2 = 0.49$.

(c) The probability that the average volume at equivalence from the technician is 100 mL is $9(0.1^2) = 0.09$.

2-154. A credit card contains 16 digits. It also contains the month and year of expiration. Suppose there are 1 million credit card holders with unique card numbers. A hacker randomly selects a 16-digit credit card number.

(a) What is the probability that it belongs to a user?
(b) Suppose a hacker has a 25% chance of correctly guessing the year your card expires and randomly selects 1 of the 12 months. What is the probability that the hacker correctly selects the month and year of expiration?

(a) $P = \dfrac{10^6}{10^{16}} = 10^{-10}$

(b) $P = 0.25\times(\dfrac{1}{12}) = 0.020833$

2-155. Eight cavities in an injection-molding tool produce plastic connectors that fall into a common stream. A sample is chosen every several minutes. Assume that the samples are independent.

(a) What is the probability that five successive samples were all produced in cavity 1 of the mold?
(b) What is the probability that five successive samples were all produced in the same cavity of the mold?
(c) What is the probability that four out of five successive samples were produced in cavity 1 of the mold?

Let A denote the event that a sample is produced in cavity one of the mold.

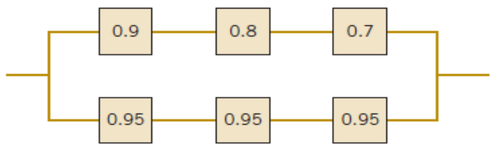
(a) By independence, $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = (\frac{1}{8})^5 = 0.00003$

(b) Let B_i be the event that all five samples are produced in cavity i. Because the B's are mutually exclusive, $P(B_1 \cup B_2 \cup ... \cup B_8) = P(B_1) + P(B_2) + ... + P(B_8)$

From part (a), $P(B_i) = (\frac{1}{8})^5$. Therefore, the answer is $8(\frac{1}{8})^5 = 0.00024$

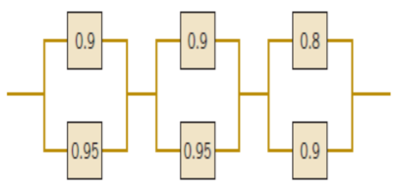
(c) By independence, $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5') = (\frac{1}{8})^4 (\frac{7}{8})$. The number of sequences in which four out of five samples are from cavity one is 5. Therefore, the answer is $5(\frac{1}{8})^4 (\frac{7}{8}) = 0.00107$.

2-156. The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



Let A denote the upper devices function. Let B denote the lower devices function.
P(A) = (0.9)(0.8)(0.7) = 0.504
P(B) = (0.95)(0.95)(0.95) = 0.8574
P(A∩B) = (0.504)(0.8574) = 0.4321
Therefore, the probability that the circuit operates = P(A∪B) = P(A) +P(B) – P(A∩B) = 0.9293

2-157. The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates?



$P = [1 – (0.1)(0.05)][1 – (0.1)(0.05)][1 – (0.2)(0.1)] = 0.9702$

2-158. Consider the endothermic reactions given below. Let *A* denote the event that a reaction's final temperature is 271 K or less. Let *B* denote the event that the heat absorbed is above target. Are these events independent?

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

P(A) = 112/204 = 0.5490, P(B) = 92/204 = 0.4510, P(A ∩ B) = 56/204 = 0.2745
Because P(A) P(B) = (0.5490)(0.4510) = 0.2476 ≠ 0.2745 = P(A ∩ B), A and B are not independent.

2-159. Consider the hospital emergency room data given below. Let *A* denote the event that a visit is to hospital 4, and let *B* denote the event that a visit results in LWBS (at any hospital). Are these events independent?

	Hospital				
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

$P(A) = 4329/22252 = 0.1945$, $P(B) = 953/22252 = 0.0428$, $P(A \cap B) = 242/22252 = 0.0109$
Because $P(A)*P(B) = (0.1945)(0.0428) = 0.0083 \neq 0.0109 = P(A \cap B)$, A and B are not independent.

2-160. Consider the well failure data given below. Let *A* denote the event that the geological formation has more than 1000 wells, and let *B* denote the event that a well failed. Are these events independent?

Geological Formation Group	Wells Failed	Wells Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

$P(A) = (1685+3733+1403)/8493 = 0.8031$, $P(B) = (170+2+443+14+29+60+46+3)/8493 = 0.0903$,
$P(A \cap B) = (170+443+60)/8493 = 0.0792$
Because $P(A)*P(B) = (0.8031)(0.0903) = 0.0725 \neq 0.0792 = P(A \cap B)$, A and B are not independent.

2-161. A Web ad can be designed from four different colors, three font types, five font sizes, three images, and five text phrases. A specific design is randomly generated by the Web server when you visit the site. Let *A* denote the event that the design color is red, and let *B* denote the event that the font size is not the smallest one. Are *A* and *B* independent events? Explain why or why not.

$P(A) = (3*5*3*5)/(4*3*5*3*5) = 0.25$, $P(B) = (4*3*4*3*5)/(4*3*5*3*5) = 0.8$,
$P(A \cap B) = (3*4*3*5) /(4*3*5*3*5) = 0.2$
Because $P(A)*P(B) = (0.25)(0.8) = 0.2 = P(A \cap B)$, A and B are independent.

2-162. Consider the code 39 is a common bar code system that consists of narrow and wide bars(black) separated by either wide or narrow spaces (white). Each character contains nine elements (five bars and four spaces). The code for a character starts and ends with a bar (either narrow or wide) and a (white) space appears between each bar. The original specification (since revised) used *exactly* two wide bars and one wide space in each character. For example, if *b* and *B* denote narrow and wide (black) bars, respectively, and *w* and *W* denote narrow and wide (white) spaces, a valid character is *bwBwBWbwb* (the number 6). Suppose that all 40 codes are equally likely (none is held back as a delimiter).
Let *A* and *B* denote the event that the first bar is wide and *B* denote the event that the second bar is wide.
Determine the following:

(a) $P(A)$ (b) $P(B)$ (c) $P(A \cap B)$ (d) Are *A* and *B* independent events?

(a) The total number of permutations of 2 wide and 3 narrow bars is $\frac{5!}{2!3!} = 10$
The number of permutations that begin with a wide bar is $\frac{4!}{1!3!} = 4$
Therefore, $P(A) = 4/10 = 0.4$
(b) A similar approach to that used in part (a) implies $P(B) = 0.4$
(c) Because a code contains exactly 2 wide bars, there is only 1 permutation with wide bars in the first and second positions. Therefore, $P(A \cap B) = 1/10 = 0.1$
(d) Because $P(A)P(B) = 0.4(0.4) = 0.16 \neq 0.1$, the events are not independent.

2-163. An integrated circuit contains 10 million logic gates (each can be a logical AND or OR circuit). Assume the probability of a gate failure is *p* and that the failures are independent. The integrated circuit fails to function if any gate fails. Determine the value for *p* so that the probability that the integrated circuit functions is 0.95.

p = probability of gate failure
A = event that the integrated circuit functions
$P(A) = 0.95 \Rightarrow P(A') = 0.05$
$(1 - p)$ = probability of gate functioning
Hence from the independence, $P(A) = (1 - p)^{10,000,000} = 0.95$.

Take logarithms to obtain $10^7\ln(1 - p) = \ln(0.95)$ and
$p = 1 - \exp[10^{-7}\ln(0.95)] = 5.13\text{x}10^{-9}$

2-164. The following table provides data on wafers categorized by location and contamination levels. Let *A* denote the event that contamination is *low*, and let *B* denote the event that the location is *center.* Are *A* and *B* independent? Why or why not?

	Location in Sputtering Tool		
Contamination	Center	Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	

A: contamination is low, B: location is center
For A and B to be independent, P(A∩B) = P(A) P(B)
P(A∩B) = 514/940 = 0.546
P(A) = 582/940 = 0.619; P(B) = 626/940 = 0.665; P(A)P(B) = 0.412. Because the probabilities are not equal, they are not independent.

2-165. The following table provides data on wafers categorized by location and contamination levels. More generally, let the number of wafers with *low* contamination from the *center* and *edge* locations be denoted as n_{lc} and n_{le}, respectively. Similarly, let n_{hc} and n_{he} denote the number of wafers with *high* contamination from the *center* and *edge* locations, respectively. Suppose that $n_{lc} = 10n_{hc}$ and $n_{le} = 10n_{he}$. That is, there are 10 times as many *low* contamination wafers as *high* ones from each location. Let *A* denote the event that contamination is low, and let *B* denote the event that the location is *center.* Are *A* and *B* independent? Does your conclusion change if the multiplier of 10 (between *low* and *high* contamination wafers) is changed from 10 to another positive integer?

	Location in Sputtering Tool		
Contamination	Center	Edge	Total
Low	514	68	582
High	112	246	358
Total	626	314	

n_{lc}=514; n_{le}=68; n_{hc}=112; n_{he}=246
$n_{lc}=10n_{hc}$; $n_{le}=10n_{he}$

	center	edge
low	$10n_{hc}$	$10n_{he}$
high	n_{hc}	n_{he}

$P(A) = (10n_{hc}+ 10n_{he})/(11n_{hc}+ 11n_{he}) = 10/11$
$P(B) = (11n_{hc})/(11n_{hc}+ 11n_{he})$
$P(A \cap B) = (10n_{hc})/(11n_{hc}+ 11n_{he})$
Because $P(A)P(B) = (10n_{hc})/(11n_{hc}+ 11n_{he})$ the events are independent.

The conclusion does not change. Even though the multiplier is changed, this relation does not change.

<u>Section 2-7</u>

2-166. Suppose that *P*(*A* | *B*) = 0.7, *P*(*A*) = 0.5, and *P*(*B*) = 0.2. Determine *P*(*B* | *A*).

Because, $P(A|B)\,P(B) = P(A \cap B) = P(B|A)\,P(A)$,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0.7(0.2)}{0.5} = 0.28$$

2-167. Suppose that *P*(*A* | *B*) = 0.4, *PA* | *B*_ = 0.2, and *P*(*B*) = 0.8. Determine *P*(*B* | *A*).

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)}=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^{'})P(B^{'})}$$

$$=\frac{0.4\times 0.8}{0.4\times 0.8+0.2\times 0.2}=0.89$$

2-168. Software to detect fraud in consumer phone cards tracks the number of metropolitan areas where calls originate each day. It is found that 1% of the legitimate users originate calls from two or more metropolitan areas in a single day. However, 30% of fraudulent users originate calls from two or more metropolitan areas in a single day. The proportion of fraudulent users is 0.01%. If the same user originates calls from two or more metropolitan areas in a single day, what is the probability that the user is fraudulent?

Let F denote a fraudulent user and let T denote a user that originates calls from two or more metropolitan areas in a day. Then,

$$P(F|T)=\frac{P(T|F)P(F)}{P(T|F)P(F)+P(T|F')P(F')}=\frac{0.30(0.0001)}{0.30(0.0001)+0.01(.9999)}=0.003$$

2-169. A new process of more accurately detecting anaerobic respiration in cells is being tested. The new process is important due to its high accuracy, its lack of extensive experimentation, and the fact that it could be used to identify five different categories of organisms: obligate anaerobes, facultative anaerobes, aerotolerant, microaerophiles, and nanaerobes instead of using a single test for each category. The process claims that it can identify obligate anaerobes with 97.8% accuracy, facultative anaerobes with 98.1% accuracy, aerotolerants with 95.9% accuracy, microaerophiles with 96.5% accuracy, and nanaerobes with 99.2% accuracy. If any category is not present, the process does not signal. Samples are prepared for the calibration of the process and 31% of them contain obligate anaerobes, 27% contain facultative anaerobes, 21% contain microaerophiles, 13% contain nanaerobes, and 8% contain aerotolerants. A test sample is selected randomly.

(a) What is the probability that the process will signal?
(b) If the test signals, what is the probability that microaerophiles are present?

(a) P=(0.31)(0.978)+(0.27)(0.981)+(0.21)(0.965)+(0.13)(0.992)+(0.08)(0.959)
= 0.97638

(b) $P=\frac{(0.21)(0.965)}{0.97638}=0.207552$

2-170. In the 2012 presidential election, exit polls from the critical state of Ohio provided the following results:

	Obama	Romney
No college degree (60%)	52%	45%
College graduate (40%)	47%	51%

If a randomly selected respondent voted for Obama, what is the probability that the person has a college degree?

Let O and C denote the respondents for Obama and the respondents with college degrees, respectively. Then
P(C | O) = P(O | C)P(C) /[P(O | C)P(C) + P(O | C')P(C')]
= 0.47(0.40)/[0.47(0.40) + 0.52(0.60)] = 0.376

2-171. Customers are used to evaluate preliminary product designs. In the past, 95% of highly successful products received good reviews, 60% of moderately successful products received good reviews, and 10% of poor products received good reviews. In addition, 40% of products have been highly successful, 35% have been moderately successful, and 25% have been poor products.
(a) What is the probability that a product attains a good review?
(b) If a new design attains a good review, what is the probability that it will be a highly successful product?
(c) If a product does not attain a good review, what is the probability that it will be a highly successful product?

Let G denote a product that received a good review. Let H, M, and P denote products that were high, moderate, and poor performers, respectively.

(a)

$$\begin{aligned} P(G) &= P(G|H)P(H) + P(G|M)P(M) + P(G|P)P(P) \\ &= 0.95(0.40) + 0.60(0.35) + 0.10(0.25) \\ &= 0.615 \end{aligned}$$

(b) Using the result from part (a)

$$P(H|G) = \frac{P(G|H)P(H)}{P(G)} = \frac{0.95(0.40)}{0.615} = 0.618$$

(c) $P(H|G') = \dfrac{P(G'|H)P(H)}{P(G')} = \dfrac{0.05(0.40)}{1 - 0.615} = 0.052$

2-172. An inspector working for a manufacturing company has a 99% chance of correctly identifying defective items and a 0.5% chance of incorrectly classifying a good item as defective. The company has evidence that 0.9% of the items its line produces are nonconforming.
(a) What is the probability that an item selected for inspection is classified as defective?
(b) If an item selected at random is classified as non defective, what is the probability that it is indeed good?

(a) $P(D)=P(D|G)P(G)+P(D|G')P(G')=(.005)(.991)+(.99)(.009)=0.013865$
(b) $P(G|D')=P(G\cap D')/P(D')=P(D'|G)P(G)/P(D')=(.995)(.991)/(1-.013865)=0.9999$

2-173. A new analytical method to detect pollutants in water is being tested. This new method of chemical analysis is important because, if adopted, it could be used to detect three different contaminants—organic pollutants, volatile solvents, and chlorinated compounds—instead of having to use a single test for each pollutant. The makers of the test claim that it can detect high levels of organic pollutants with 99.7% accuracy, volatile solvents with 99.95% accuracy, and chlorinated compounds with 89.7% accuracy. If a pollutant is not present, the test does not signal. Samples are prepared for the calibration of the test and 60% of them are contaminated with organic pollutants, 27% with volatile solvents, and 13% with traces of chlorinated compounds. A test sample is selected randomly.
(a) What is the probability that the test will signal?
(b) If the test signals, what is the probability that chlorinated compounds are present?

Denote as follows: S = signal, O = organic pollutants, V = volatile solvents, C = chlorinated compounds
(a) $P(S) = P(S|O)P(O)+P(S|V)P(V)+P(S|C)P(C) = 0.997(0.60) + 0.9995(0.27) + 0.897(0.13) = 0.9847$
(b) $P(C|S) = P(S|C)P(C)/P(S) = (0.897)(0.13)/0.9847 = 0.1184$

2-174. Consider the endothermic reactions given below. Use Bayes' theorem to calculate the probability that a reaction's final temperature is 271 K or less given that the heat absorbed is above target.

Final Temperature Conditions	Heat Absorbed (cal)	
	Below Target	Above Target
266 K	12	40
271 K	44	16
274 K	56	36

Let A denote the event that a reaction final temperature is 271 K or less
Let B denote the event that the heat absorbed is above target

$$\begin{aligned} P(A|B) = \frac{P(A\cap B)}{P(B)} &= \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A')P(B|A')} \\ &= \frac{(0.5490)(0.5)}{(0.5490)(0.5)+(0.4510)(0.3913)} = 0.6087 \end{aligned}$$

2-175. Consider the hospital emergency room data given below. Use Bayes' theorem to calculate the probability that a person visits hospital 4 given they are LWBS.

	Hospital				
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

Let L denote the event that a person is LWBS
Let A denote the event that a person visits Hospital 1
Let B denote the event that a person visits Hospital 2
Let C denote the event that a person visits Hospital 3
Let D denote the event that a person visits Hospital 4

$$P(D\mid L)=\frac{P(L\mid D)P(D)}{P(L\mid A)P(A)+P(L\mid B)P(B)+P(L\mid C)P(C)+P(L\mid D)P(D)}$$
$$=\frac{(0.0559)(0.1945)}{(0.0368)(0.2378)+(0.0386)(0.3142)+(0.0436)(0.2535)+(0.0559)(0.1945)}$$
$$=0.2540$$

2-176. Consider the well failure data given below. Use Bayes' theorem to calculate the probability that a randomly selected well is in the gneiss group given that the well has failed.

	Wells	
Geological Formation Group	Failed	Total
Gneiss	170	1685
Granite	2	28
Loch raven schist	443	3733
Mafic	14	363
Marble	29	309
Prettyboy schist	60	1403
Other schists	46	933
Serpentine	3	39

Let A denote the event that a well is failed
Let B denote the event that a well is in Gneiss
Let C denote the event that a well is in Granite
Let D denote the event that a well is in Loch raven schist
Let E denote the event that a well is in Mafic
Let F denote the event that a well is in Marble
Let G denote the event that a well is in Prettyboy schist
Let H denote the event that a well is in Other schist
Let I denote the event that a well is in Serpentine

$$P(B\mid A)=\frac{P(A\mid B)P(B)}{P(A\mid B)P(B)+P(A\mid C)P(C)+P(A\mid D)P(D)+P(A\mid E)P(E)+P(A\mid F)P(F)+P(A\mid G)P(E)+P(A\mid G)P(G)+P(A\mid H)P(H)}$$
$$=\frac{\left(\frac{170}{1685}\right)\left(\frac{1685}{8493}\right)}{\left(\frac{170}{1685}\right)\left(\frac{1685}{8493}\right)+\left(\frac{2}{28}\right)\left(\frac{28}{8493}\right)+\left(\frac{443}{3733}\right)\left(\frac{3733}{8493}\right)+\left(\frac{14}{363}\right)\left(\frac{363}{8493}\right)+\left(\frac{29}{309}\right)\left(\frac{309}{8493}\right)+\left(\frac{60}{1403}\right)\left(\frac{1403}{8493}\right)+\left(\frac{46}{933}\right)\left(\frac{933}{8493}\right)+\left(\frac{3}{39}\right)\left(\frac{39}{8493}\right)}$$
$$=0.2216$$

2-177. Two Web colors are used for a site advertisement. If a site visitor arrives from an affiliate, the probabilities of the blue or green colors being used in the advertisement are 0.8 and 0.2, respectively. If the site visitor arrives from a search site, the probabilities of blue and green colors in the advertisement are 0.4 and 0.6, respectively. The proportions of visitors from affiliates and search sites are 0.3 and 0.7, respectively. What is the probability that a visitor is from a search site given that the blue ad was viewed?

Denote as follows: A = affiliate site, S = search site, B =blue, G =green

$$P(S\mid B)=\frac{P(B\mid S)P(S)}{P(B\mid S)P(S)+P(B\mid A)P(A)}=\frac{(0.4)(0.7)}{(0.4)(0.7)+(0.8)(0.3)}=0.5$$

2-178. The article ["Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis," *Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381– 3390)] considered four treatment groups. The groups consisted of patients with different drug therapies (such as prednisone and infliximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or initial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease progression. The number of patients without progression of joint damage was 76 of 114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose that a patient is selected randomly. Let *A* denote the event that the patient is in group 1, and let *B* denote the event that there is no progression.
Determine the following probabilities:

(a) $P(B)$ (b) $P(B \mid A)$ (c) $P(A \mid B)$

(a) P(B) = P(B|G1) P(G1)+P(B|G2) P(G2)+P(B|G3) P(G3)+P(P|G4) P(G4) = 0.802

(b) P(B|A) = $\frac{P(A \cap B)}{P(A)}$ = 76/114 = 0.667

(c) P(A|B) =

$$\frac{P(B|A)P(A)}{\mathrm{P(B|G1)P(G1)+P(B|G2)P(G2)+P(B|G3)P(G3)+P(P|G4)P(G4)}} = \frac{0.667(0.244)}{0.802} = 0.203$$

2-179. An e-mail filter is planned to separate valid e-mails from spam. The word *free* occurs in 60% of the spam messages and only 4% of the valid messages. Also, 20% of the messages are spam. Determine the following probabilities:
(a) The message contains *free.*
(b) The message is spam given that it contains *free.*
(c) The message is valid given that it does not contain *free.*

F: Free; S: Spam; V: Valid

P(F|S) = 0.6, P(F|V) = 0.04

(a) P(F) = P(F|S) P(S)+P(F|V) P(V) = 0.6(0.2) + 0.04(0.8) = 0.152

(b) P(S|F) = $\frac{P(F|S)P(S)}{P(F)} = \frac{0.6(0.2)}{0.152}$ = 0.789

(c) P(V|F') = $\frac{P(F'|V)P(V)}{P(F')} = \frac{(0.96)0.8}{1-0.152} = 0.906$

2-180. A recreational equipment supplier finds that among orders that include tents, 40% also include sleeping mats. Only 5% of orders that do not include tents do include sleeping mats. Also, 20% of orders include tents. Determine the following probabilities:
(a) The order includes sleeping mats.
(b) The order includes a tent given it includes sleeping mats.

SM: Sleeping Mats; T:Tents;

P(SM|T) = 0.4; P(SM|T') = 0.05; P(T) = 0.2

(a) P(SM) = P(SM|T)P(T) + P(SM|T')P(T') = 0.4(0.2) + 0.05(0.8) = 0.12

(b) P(T|SM) = $\frac{\mathrm{P(SM|T)P(T)}}{P(SM)} = \frac{0.4X0.2}{0.12} = 0.667$

2-181. The probabilities of poor print quality given no printer problem, misaligned paper, high ink viscosity, or printer-head debris are 0, 0.3, 0.4, and 0.6, respectively. The probabilities of no printer problem, misaligned paper, high ink viscosity, or printer-head debris are 0.8, 0.02, 0.08, and 0.1, respectively.
(a) Determine the probability of high ink viscosity given poor print quality.
(b) Given poor print quality, what problem is most likely?

NP = no problem; PP = poor print; MP = misaligned paper

HV = high ink viscosity; HD = print head debris

P(MP) = 0.02; P(HV) = 0.08; P(HD) = 0.1; P(NP) = 0.8

P(PP|NP) = 0; P(PP|MP) = 0.3; P(PP|HV) = 0.4; P(PP|HD) = 0.6; P(NP) = 0.8

(a) P(HV|PP)= $\frac{P(PP|HV)P(HV)}{P(PP)}$

$P(PP) = P(PP|HV)P(HV) + P(PP|NP)P(NP) + P(PP|MP)P(MP) + P(PP|HD)P(HD)$

$= 0.4(0.08) + 0(0.8) + 0.3(0.02) + 0.6(0.1) = 0.98$

Therefore, P(HV|PP) = $\frac{0.4(0.08)}{0.098} = \frac{0.032}{0.098} = 0.327$

(b) P(HV|PP) = P(PP|HV)P(HV)/P(PP) = 0.032/0.098 = 0.327

P(NP|PP) = P(PP|NP)P(NP)/P(PP) = 0

P(MP|PP)= P(PP|MP)P(MP)/P(PP) = 0.006/0.098 = 0.061

P(HD|PP) = P(PP | HD)P(HD)/P(PP) = 0.06/0.098 = 0.612

The problem most likely given poor print quality is head debris.

Section 2-8

2-182. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The time until a projectile returns to earth.
(b) The number of times a transistor in a computer memory changes state in one operation.
(c) The volume of gasoline that is lost to evaporation during the filling of a gas tank.
(d) The outside diameter of a machined shaft.

(a) continuous (b) discrete (c) continuous (d) continuous

2-183. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The number of cracks exceeding one-half inch in 10 miles of an interstate highway.
(b) The weight of an injection-molded plastic part.
(c) The number of molecules in a sample of gas.
(d) The concentration of output from a reactor.
(e) The current in an electronic circuit.

(a) discrete (b) continuous (c) discrete, but large values might be modeled as continuous
(d) continuous (e) continuous

2-184. Decide whether a discrete or continuous random variable is the best model for each of the following variables:
(a) The time for a computer algorithm to assign an image to a category.
(b) The number of bytes used to store a file in a computer.
(c) The ozone concentration in micrograms per cubic meter.
(d) The ejection fraction (volumetric fraction of blood pumped from a heart ventricle with each beat).
(e) The fluid flow rate in liters per minute.

(a) continuous (b) discrete, but large values might be modeled as continuous
(c) continuous (d) continuous (e) continuous

Supplemental Exercises

2-185. Samples of laboratory glass are in small, light packaging or heavy, large packaging. Suppose that 2% and 1%, respectively, of the sample shipped in small and large packages, respectively, break during transit. If 60% of the samples are shipped in large packages and 40% are shipped in small packages, what proportion of samples break during shipment?

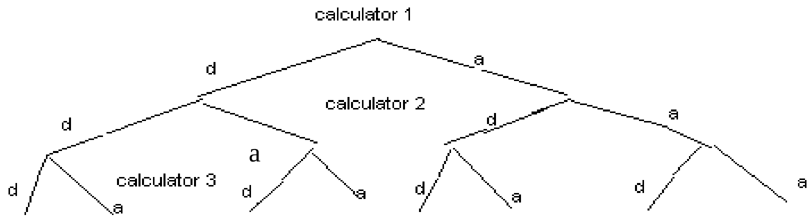
Let B denote the event that a glass breaks.
Let L denote the event that large packaging is used.
P(B)= P(B|L)P(L) + P(B|L')P(L')
= 0.01(0.60) + 0.02(0.40) = 0.014

2-186. A sample of three calculators is selected from a manufacturing line, and each calculator is classified as either defective or acceptable. Let *A*, *B*, and *C* denote the events that the first, second, and third calculators, respectively, are defective.
(a) Describe the sample space for this experiment with a tree diagram.

Use the tree diagram to describe each of the following events:

(b) A (c) B (d) $A \cap B$ (e) $B \cup C$

Let "d" denote a defective calculator and let "a" denote an acceptable calculator



(a) $S = \{ddd, add, dda, ada, dad, aad, daa, aaa\}$
(b) $A = \{ddd, dda, dad, daa\}$
(c) $B = \{ddd, dda, add, ada\}$
(d) $A \cap B = \{ddd, dda\}$
(e) $B \cup C = \{ddd, dda, add, ada, dad, aad\}$

2-187. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and edge finish. The results of 100 parts are summarized as follows:

		Edge Finish	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

Let A denote the event that a sample has excellent surface finish, and let B denote the event that a sample has excellent edge finish. If a part is selected at random, determine the following probabilities:

(a) $P(A)$ (b) $P(B)$ (c) $P(A')$ (d) $P(A \cap B)$ (e) $P(A \cup B)$ (f) $P(A' \cup B)$

Let A = excellent surface finish; B = excellent length
(a) P(A) = 82/100 = 0.82
(b) P(B) = 90/100 = 0.90
(c) P(A') = 1 – 0.82 = 0.18
(d) P(A∩B) = 80/100 = 0.80
(e) P(A∪B) = 0.92
(f) P(A'∪B) = 0.98

2-188. Shafts are classified in terms of the machine tool that was used for manufacturing the shaft and conformance to surface finish and roundness.

Tool 1		Roundness Conforms	
		Yes	No
Surface Finish	Yes	200	1
Conforms	No	4	2

Tool 2		Roundness Conforms	
		Yes	No
Surface Finish	Yes	145	4
Conforms	No	8	6

(a) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or to roundness requirements or is from tool 1?
(b) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or does not conform to roundness requirements or is from tool 2?
(c) If a shaft is selected at random, what is the probability that the shaft conforms to both surface finish and roundness requirements or the shaft is from tool 2?

(d) If a shaft is selected at random, what is the probability that the shaft conforms to surface finish requirements or the shaft is from tool 2?

(a) (207+350+357-201-204-345+200)/370 = 0.9838
(b) 366/370 = 0.989
(c) (200+163)/370 = 363/370 = 0.981
(d) (201+163)/370 = 364/370 = 0.984

2-189. If *A*, *B*, and *C* are mutually exclusive events, is it possible for $P(A) = 0.3$, $P(B) = 0.4$, and $P(C) = 0.5$? Why or why not?

If A,B,C are mutually exclusive, then $P(A \cup B \cup C) = P(A) + P(B) + P(C) = 0.3 + 0.4 + 0.5 =$ 1.2, which greater than 1. Therefore, P(A), P(B),and P(C) cannot equal the given values.

2-190. The analysis of shafts for a compressor is summarized by conformance to specifications:

		Roundness Conforms	
		Yes	No
Surface finish	Yes	345	5
Conforms	No	12	8

(a) If we know that a shaft conforms to roundness requirements, what is the probability that it conforms to surface finish requirements?
(b) If we know that a shaft does not conform to roundness requirements, what is the probability that it conforms to surface finish requirements?

(a) 345/357 (b) 5/13

2-191. A researcher receives 100 containers of oxygen. Of those containers, 20 have oxygen that is not ionized, and the rest are ionized. Two samples are randomly selected, without replacement, from the lot.
(a) What is the probability that the first one selected is not ionized?
(b) What is the probability that the second one selected is not ionized given that the first one was ionized?
(c) What is the probability that both are ionized?
(d) How does the answer in part (b) change if samples selected were replaced prior to the next selection?

(a) P(the first one selected is not ionized)=20/100=0.2
(b) P(the second is not ionized given the first one was ionized) =20/99=0.202
(c) P(both are ionized)
= P(the first one selected is ionized) $\times$ P(the second is ionized given the first one was ionized)
= (80/100)$\times$(79/99)=0.638
(d) If samples selected were replaced prior to the next selection,
P(the second is not ionized given the first one was ionized) =20/100=0.2.
The event of the first selection and the event of the second selection are independent.

2-192. A lot contains 15 castings from a local supplier and 25 castings from a supplier in the next state. Two castings are selected randomly, without replacement, from the lot of 40. Let *A* be the event that the first casting selected is from the local supplier, and let *B* denote the event that the second casting is selected from the local supplier. Determine:

(a) $P(A)$ (b) $P(B \mid A)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

Suppose that 3 castings are selected at random, without replacement, from the lot of 40. In addition to the definitions of events *A* and *B*, let *C* denote the event that the third casting selected is from the local supplier. Determine:

(e) $P(A \cap B \cap C)$ (f) $P(A \cap B \cap C')$

(a) P(A) = 15/40
(b) $P(B|A)$ = 14/39
(c) $P(A \cap B)$ = P(A) P(B/A) = (15/40) (14/39) = 0.135
(d) $P(A \cup B)$ = 1 – P(A' and B') = $1-\left(\frac{25}{40}\right)\left(\frac{24}{39}\right)=0.615$
A = first is local, B = second is local, C = third is local
(e) $P(A \cap B \cap C)$ = (15/40)(14/39)(13/38) = 0.046

(f) P(A ∩ B ∩ C') = (15/40)(14/39)(25/39) = 0.089

2-193. In the manufacturing of a chemical adhesive, 3% of all batches have raw materials from two different lots. This occurs when holding tanks are replenished and the remaining portion of a lot is insufficient to fill the tanks. Only 5% of batches with material from a single lot require reprocessing. However, the viscosity of batches consisting of two or more lots of material is more difficult to control, and 40% of such batches require additional processing to achieve the required viscosity.

Let *A* denote the event that a batch is formed from two different lots, and let *B* denote the event that a lot requires additional processing. Determine the following probabilities:

(a) $P(A)$ (b) $P(A')$ (c) $P(B \mid A)$ (d) $P(B \mid A')$
(e) $P(A \cap B)$ (f) $P(A \cap B')$ (g) $P(B)$

(a) P(A) = 0.03
(b) P(A') = 0.97
(c) P(B|A) = 0.40
(d) P(B|A') = 0.05

(e) $P(A \cap B) = P(B|A)P(A) = (0.40)(0.03) = 0.012$

(f) $P(A \cap B') = P(B'|A)P(A) = (0.60)(0.03) = 0.018$

(g) $P(B) = P(B|A)P(A) + P(B|A')P(A') = (0.40)(0.03) + (0.05)(0.97) = 0.0605$

2-194. Incoming calls to a customer service center are classified as complaints (75% of calls) or requests for information (25% of calls). Of the complaints, 40% deal with computer equipment that does not respond and 57% deal with incomplete software installation; in the remaining 3% of complaints, the user has improperly followed the installation instructions. The requests for information are evenly divided on technical questions (50%) and requests to purchase more products (50%).
(a) What is the probability that an incoming call to the customer service center will be from a customer who has not followed installation instructions properly?
(b) Find the probability that an incoming call is a request for purchasing more products.

Let U denote the event that the user has improperly followed installation instructions.
Let C denote the event that the incoming call is a complaint.
Let P denote the event that the incoming call is a request to purchase more products.
Let R denote the event that the incoming call is a request for information.
a) P(U|C)P(C) = (0.75)(0.03) = 0.0225
b) P(P|R)P(R) = (0.50)(0.25) = 0.125

2-195. A congested computer network has a 0.002 probability of losing a data packet, and packet losses are independent events. A lost packet must be resent.
(a) What is the probability that an e-mail message with 100 packets will need to be resent?
(b) What is the probability that an e-mail message with 3 packets will need exactly 1 to be resent?
(c) If 10 e-mail messages are sent, each with 100 packets, what is the probability that at least 1 message will need some packets to be resent?

(a) $P = 1-(1-0.002)^{100} = 0.18143$

(b) $P = C_3^1(0.998^2)0.002 = 0.005976$

(c) $P = 1-[(1-0.002)^{100}]^{10} = 0.86494$

2-196. Samples of a cast aluminum part are classified on the basis of surface finish (in microinches) and length measurements. The results of 100 parts are summarized as follows:

		Length	
		Excellent	Good
Surface	Excellent	80	2
Finish	Good	10	8

Let *A* denote the event that a sample has excellent surface finish, and let *B* denote the event that a sample has excellent length. Are events *A* and *B* independent?

$P(A \cap B) = 80/100$, $P(A) = 82/100$, $P(B) = 90/100$.

Then, $P(A \cap B) \neq P(A)P(B)$, so A and B are <u>not</u> independent.

2-197. An optical storage device uses an error recovery procedure that requires an immediate satisfactory readback of any written data. If the readback is not successful after three writing operations, that sector of the disk is eliminated as unacceptable for data storage. On an acceptable portion of the disk, the probability of a satisfactory readback is 0.98. Assume the readbacks are independent. What is the probability that an acceptable portion of the disk is eliminated as unacceptable fordata storage?

Let A_i denote the event that the ith readback is successful. By independence,

$$P(A_1' \cap A_2' \cap A_3') = P(A_1')P(A_2')P(A_3') = (0.02)^3 = 0.000008.$$

2-198. Semiconductor lasers used in optical storage products require higher power levels for write operations than for read operations. High-power-level operations lower the useful life of the laser. Lasers in products used for backup of higher-speed magnetic disks primarily write, and the probability that the useful life exceeds five years is 0.95. Lasers that are in products that are used for main storage spend approximately an equal amount of time reading and writing, and the probability that the useful life exceeds five years is 0.995. Now, 25% of the products from a manufacturer are used for backup and 75% of the products are used for main storage.

Let *A* denote the event that a laser's useful life exceeds five years, and let *B* denote the event that a laser is in a product that is used for backup.

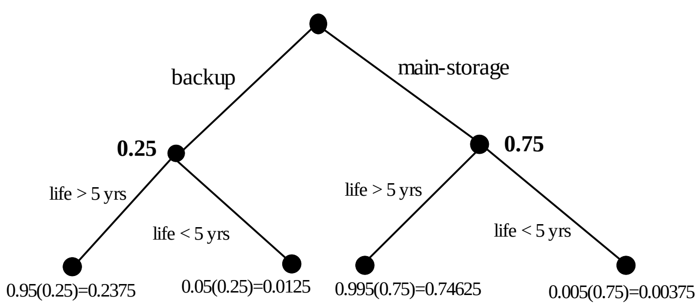
Use a tree diagram to determine the following:

(a) $P(B)$ (b) $P(A \mid B)$ (c) $P(A \mid B')$

(d) $P(A \cap B)$ (e) $P(A \cap B')$ (f) $P(A)$

(g) What is the probability that the useful life of a laser exceeds five years?

(h) What is the probability that a laser that failed before five years came from a product used for backup?



(a) $P(B) = 0.25$

(b) $P(A|B) = 0.95$

(c) $P(A|B') = 0.995$

(d) $P(A \cap B) = P(A|B)P(B) = 0.95(0.25) = 0.2375$

(e) $P(A \cap B') = P(A|B')P(B') = 0.995(0.75) = 0.74625$

(f) $P(A) = P(A \cap B) + P(A \cap B') = 0.95(0.25) + 0.995(0.75) = 0.98375$

(g) $0.95(0.25) + 0.995(0.75) = 0.98375$.

(h)

$$P(B|A') = \frac{P(A'|B)P(B)}{P(A'|B)P(B) + P(A'|B')P(B')} = \frac{0.05(0.25)}{0.05(0.25) + 0.005(0.75)} = 0.769$$

2-199. Energy released from cells breaks the molecular bond and converts ATP (adenosine triphosphate) into ADP (adenosine diphosphate). Storage of ATP in muscle cells (even for an athlete) can sustain maximal muscle power only for less than five seconds (a short dash). Three systems are used to replenish ATP—phosphagen system, glycogen-lactic acid system (anaerobic), and aerobic respiration—but the first is useful only for less than 10 seconds, and even the second system provides less than two minutes of ATP. An endurance athlete needs to perform below the anaerobic threshold to sustain energy for extended periods. A sample of 100 individuals is described by the energy system used in exercise at different intensity levels.

	Primarily Aerobic	
Period	Yes	No
1	50	7
2	13	30

Let *A* denote the event that an individual is in period 2, and let *B* denote the event that the energy is primarily aerobic. Determine the number of individuals in

(a) $A' \cap B$ (b) B' (c) $A \cup B$

(a) $A' \cap B = 50$
(b) B'=37
(c) $A \cup B = 93$

2-200. A sample preparation for a chemical measurement is completed correctly by 25% of the lab technicians, completed with a minor error by 70%, and completed with a major error by 5%.
(a) If a technician is selected randomly to complete the preparation, what is the probability that it is completed without error?
(b) What is the probability that it is completed with either a minor or a major error?

(a) 0.25
(b) 0.75

2-201. In circuit testing of printed circuit boards, each board either fails or does not fail the test. A board that fails the test is then checked further to determine which one of five defect types is the primary failure mode. Represent the sample space for this experiment.

Let D_i denote the event that the primary failure mode is type i and let A denote the event that a board passes the test. The sample space is S = $\{A, A'D_1, A'D_2, A'D_3, A'D_4, A'D_5\}$.

2-202. The data from 200 machined parts are summarized as follows:

	Depth of Bore	
Edge Condition	Above Target	Below Target
Coarse	15	10
Moderate	25	20
Smooth	50	80

(a) What is the probability that a part selected has a moderate edge condition and a below-target bore depth?
(b) What is the probability that a part selected has a moderate edge condition or a below-target bore depth?
(c) What is the probability that a part selected does not have a moderate edge condition or does not have a below-target bore depth?

(a) 20/200 (b) 135/200 (c) 65/200

2-203. Computers in a shipment of 100 units contain a portable hard drive, solid-state memory, or both, according to the following table:

	Portable Hard Drive	
Solid-state memory	Yes	No
Yes	15	80
No	4	1

Let *A* denote the event that a computer has a portable hard drive, and let *B* denote the event that a computer has a solidstate memory. If one computer is selected randomly, compute

(a) $P(A)$ (b) $P(A \cap B)$ (c) $P(A \cup B)$ (d) $P(A' \cap B)$ (e) $P(A \mid B)$

(a) P(A) = 19/100 = 0.19
(b) $P(A \cap B) = 15/100 = 0.15$
(c) $P(A \cup B) = (19 + 95 - 15)/100 = 0.99$
(d) $P(A' \cap B) = 80/100 = 0.80$
(e) $P(A|B) = P(A \cap B)/P(B) = 0.158$

2-204. The probability that a customer's order is not shipped on time is 0.05. A particular customer places three orders, and the orders are placed far enough apart in time that they can be considered to be independent events.
(a) What is the probability that all are shipped on time?
(b) What is the probability that exactly one is not shipped on time?
(c) What is the probability that two or more orders are not shipped on time?

Let A_i denote the event that the *i*th order is shipped on time.

(a) By independence, $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) = (0.95)^3 = 0.857$

(b) Let

$$B_1 = A_1' \cap A_2 \cap A_3$$
$$B_2 = A_1 \cap A_2' \cap A_3$$
$$B_3 = A_1 \cap A_2 \cap A_3'$$

Then, because the B's are mutually exclusive,

$$\begin{aligned} P(B_1 \cup B_2 \cup B_3) &= P(B_1) + P(B_2) + P(B_3) \\ &= 3(0.95)^2(0.05) \\ &= 0.135 \end{aligned}$$

(c) Let

$$B_1 = A_1' \cap A_2' \cap A_3$$
$$B_2 = A_1' \cap A_2 \cap A_3'$$
$$B_3 = A_1 \cap A_2' \cap A_3'$$
$$B_4 = A_1' \cap A_2' \cap A_3'$$

Because the B's are mutually exclusive,

$$\begin{aligned} P(B_1 \cup B_2 \cup B_3 \cup B_4) &= P(B_1) + P(B_2) + P(B_3) + P(B_4) \\ &= 3(0.05)^2(0.95) + (0.05)^3 \\ &= 0.00725 \end{aligned}$$

2-205. Let E_1, E_2, and E_3 denote the samples that conform to a percentage of solids specification, a molecular weight specification, and a color specification, respectively. A total of 240 samples are classified by the E_1, E_2, and E_3 specifications, where *yes* indicates that the sample conforms.

E_3 yes

		E_2		
		Yes	No	Total
E_1	Yes	200	1	201
	No	5	4	9
Total		205	5	210

E_3 no

		E_2		
		Yes	No	Total
E_1	Yes	20	4	24
	No	6	0	6
Total		26	4	30

(a) Are E_1, E_2, and E_3 mutually exclusive events?
(b) Are E_1', E_2' , and E_3' mutually exclusive events?
(c) What is $P(E_1'$ or E_2' or $E_3')$?
(d) What is the probability that a sample conforms to all three specifications?
(e) What is the probability that a sample conforms to the E_1 or E_3 specification?
(f) What is the probability that a sample conforms to the E_1 or E_2 or E_3 specification?

(a) No, $P(E_1 \cap E_2 \cap E_3) \neq 0$
(b) No, $E_1' \cap E_2'$ is not $\varnothing$

(c) $P(E_1' \cup E_2' \cup E_3') = P(E_1') + P(E_2') + P(E_3') – P(E_1' \cap E_2') - P(E_1' \cap E_3') - P(E_2' \cap E_3') + P(E_1' \cap E_2' \cap E_3') = 40/240$

(d) $P(E_1 \cap E_2 \cap E_3) = 200/240$

(e) $P(E_1 \cup E_3) = P(E_1) + P(E_3) – P(E_1 \cap E_3) = 234/240$

(f) $P(E_1 \cup E_2 \cup E_3) = 1 – P(E_1' \cap E_2' \cap E_3') = 1 - 0 = 1$

2-206. Transactions to a computer database are either new items or changes to previous items. The addition of an item can be completed in less than 100 milliseconds 90% of the time, but only 20% of changes to a previous item can be completed in less than this time. If 30% of transactions are changes, what is the probability that a transaction can be completed in less than 100 milliseconds?

(a) (0.20)(0.30) +(0.7)(0.9) = 0.69

2-207. A steel plate contains 20 bolts. Assume that 5 bolts are not torqued to the proper limit. 4 bolts are selected at random, without replacement, to be checked for torque.

(a) What is the probability that all 4 of the selected bolts are torqued to the proper limit?

(b) What is the probability that at least 1 of the selected bolts is *not* torqued to the proper limit?

Let A_i denote the event that the *i*th bolt selected is not torqued to the proper limit.

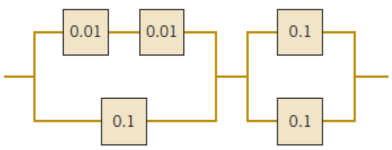
(a) Then,

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_4 | A_1 \cap A_2 \cap A_3) P(A_1 \cap A_2 \cap A_3)$$
$$= P(A_4 | A_1 \cap A_2 \cap A_3) P(A_3 | A_1 \cap A_2) P(A_2 | A_1) P(A_1)$$
$$= \left(\frac{12}{17}\right)\left(\frac{13}{18}\right)\left(\frac{14}{19}\right)\left(\frac{15}{20}\right) = 0.282$$

(b) Let B denote the event that at least one of the selected bolts are not properly torqued. Thus, B' is the event that all bolts are properly torqued. Then,

$$P(B) = 1 - P(B') = 1 - \left(\frac{15}{20}\right)\left(\frac{14}{19}\right)\left(\frac{13}{18}\right)\left(\frac{12}{17}\right) = 0.718$$

2-208. The following circuit operates if and only if there is a path of functional devices from left to right. Assume devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit operates?



Let A,B denote the event that the first, second portion of the circuit operates.

Then, P(A) = (0.99)(0.99)+0.9-(0.99)(0.99)(0.9) = 0.998

P(B) = 0.9+0.9-(0.9)(0.9) = 0.99 and

$P(A \cap B) = P(A) P(B) = (0.998) (0.99) = 0.988$

2-209. The probability that concert tickets are available by telephone is 0.92. For the same event, the probability that tickets are available through a Web site is 0.95. Assume that these two ways to buy tickets are independent. What is the probability that someone who tries to buy tickets through the Web and by phone will obtain tickets?

A_1 = by telephone, A_2 = website; $P(A_1) = 0.92$, $P(A_2) = 0.95$;

By independence $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0.92 + 0.95 - 0.92(0.95) = 0.996$

2-210. The British government has stepped up its information campaign regarding foot-and-mouth disease by mailing brochures to farmers around the country. It is estimated that 99% of Scottish farmers who receive the brochure possess enough information to deal with an outbreak of the disease, but only 90% of those without the brochure can deal with an outbreak. After the first three months of mailing, 95% of the farmers in Scotland had received the informative brochure. Compute the probability that a randomly selected farmer will have enough information to deal effectively with an outbreak of the disease.

P(Possess) = 0.95(0.99) +(0.05)(0.90) = 0.9855

2-211. In an automated filling operation, the probability of an incorrect fill when the process is operated at a low speed is 0.001. When the process is operated at a high speed, the probability of an incorrect fill is 0.01. Assume that 30% of the containers are filled when the process is operated at a high speed and the remainder are filled when the process is operated at a low speed.

(a) What is the probability of an incorrectly filled container?

(b) If an incorrectly filled container is found, what is the probability that it was filled during the high-speed operation?

Let D denote the event that a container is incorrectly filled and let H denote the event that a container is filled under high-speed operation. Then,

(a) $P(D) = P(D|H)P(H) + P(D|H')P(H') = 0.01(0.30) + 0.001(0.70) = 0.0037$

(b) $P(H|D) = \dfrac{P(D|H)P(H)}{P(D)} = \dfrac{0.01(0.30)}{0.0037} = 0.8108$

2-212. An encryption-decryption system consists of three elements: encode, transmit, and decode. A faulty encode occurs in 0.5% of the messages processed, transmission errors occur in 1% of the messages, and a decode error occurs in 0.1% of the messages. Assume the errors are independent.

(a) What is the probability of a completely defect-free message?

(b) What is the probability of a message that has either an encode or a decode error?

(a) $P(E' \cap T' \cap D') = (0.995)(0.99)(0.999) = 0.984$

(b) $P(E \cup D) = P(E) + P(D) - P(E \cap D) = 0.005995$

2-213. It is known that two defective copies of a commercial software program were erroneously sent to a shipping lot that now has a total of 75 copies of the program. A sample of copies will be selected from the lot without replacement.

(a) If three copies of the software are inspected, determine the probability that exactly one of the defective copies will be found.

(b) If three copies of the software are inspected, determine the probability that both defective copies will be found.

(c) If 73 copies are inspected, determine the probability that both copies will be found. (*Hint*: Work with the copies that remain in the lot.)

D = defective copy

(a) $P(D = 1) = \left(\frac{2}{75}\right)\left(\frac{73}{74}\right)\left(\frac{72}{73}\right) + \left(\frac{73}{75}\right)\left(\frac{2}{74}\right)\left(\frac{72}{73}\right) + \left(\frac{73}{75}\right)\left(\frac{72}{74}\right)\left(\frac{2}{73}\right) = 0.0778$

(b) $P(D = 2) = \left(\frac{2}{75}\right)\left(\frac{1}{74}\right)\left(\frac{73}{73}\right) + \left(\frac{2}{75}\right)\left(\frac{73}{74}\right)\left(\frac{1}{73}\right) + \left(\frac{73}{75}\right)\left(\frac{2}{74}\right)\left(\frac{1}{73}\right) = 0.00108$

(c) Let A represent the event that the two items NOT inspected are not defective. Then, P(A)=(73/75)(72/74)=0.947.

2-214. A robotic insertion tool contains 10 primary components. The probability that any component fails during the warranty period is 0.01. Assume that the components fail independently and that the tool fails if any component fails. What is the probability that the tool fails during the warranty period?

The tool fails if any component fails. Let F denote the event that the tool fails. Then, $P(F') = 0.99^{10}$ by independence and $P(F) = 1 - 0.99^{10} = 0.0956$

2-215. An e-mail message can travel through one of two server routes. The probability of transmission error in each of the servers and the proportion of messages that travel each route are shown in the following table. Assume that the servers are independent.

		Probability of Error			
	Percentage of Messages	Server 1	Server 2	Server 3	Server 4
Route 1	30	0.01	0.015	—	—
Route 2	70	—	—	0.02	0.003

(a) What is the probability that a message will arrive without error?

(b) If a message arrives in error, what is the probability it was sent through route 1?

(a) (0.3)(0.99)(0.985) + (0.7)(0.98)(0.997) = 0.9764

(b) $P(route1|E) = \dfrac{P(E|route1)P(route1)}{P(E)} = \dfrac{0.02485(0.30)}{1-0.9764} = 0.3159$

2-216. A machine tool is idle 15% of the time. You request immediate use of the tool on five different occasions during the year. Assume that your requests represent independent events.
(a) What is the probability that the tool is idle at the time of all of your requests?
(b) What is the probability that the machine is idle at the time of exactly four of your requests?
(c) What is the probability that the tool is idle at the time of at least three of your requests?

(a) By independence, $0.15^5 = 7.59 \times 10^{-5}$

(b) Let A_i denote the events that the machine is idle at the time of your ith request. Using independence, the requested probability is

$$P(A_1A_2A_3A_4A_5' \;\; or \;\; A_1A_2A_3A_4'A_5 \;\; or \;\; A_1A_2A_3'A_4A_5 \;\; or \;\; A_1A_2'A_3A_4A_5 \;\; or \;\; A_1'A_2A_3A_4A_5)$$
$$= 5(0.15^4)(0.85^1)$$
$$= 0.00215$$

(c) As in part b, the probability of 3 of the events is

$$P(A_1A_2A_3A_4'A_5' \;\; or \;\; A_1A_2A_3'A_4A_5' \;\; or \;\; A_1A_2A_3'A_4'A_5 \;\; or \;\; A_1A_2'A_3A_4A_5' \;\; or \;\; A_1A_2'A_3A_4'A_5 \; or$$
$$A_1A_2'A_3'A_4A_5 \;\; or \;\; A_1'A_2A_3A_4A_5' \;\; or \;\; A_1'A_2A_3A_4'A_5 \;\; or \;\; A_1'A_2A_3'A_4A_5 \;\; or \;\; A_1'A_2'A_3A_4A_5)$$
$$= 10(0.15^3)(0.85^2)$$
$$= 0.0244$$

For the probability of at least 3, add answer parts a) and b) to the above to obtain the requested probability. Therefore, the answer is 0.0000759 + 0.0022 + 0.0244 = 0.0267

2-217. A lot of 50 spacing washers contains 30 washers that are thicker than the target dimension. Suppose that 3 washers are selected at random, without replacement, from the lot.
(a) What is the probability that all 3 washers are thicker than the target?
(b) What is the probability that the third washer selected is thicker than the target if the first 2 washers selected are thinner than the target?
(c) What is the probability that the third washer selected is thicker than the target?

Let A_i denote the event that the ith washer selected is thicker than target.

(a) $\left(\dfrac{30}{50}\right)\left(\dfrac{29}{49}\right)\left(\dfrac{28}{48}\right) = 0.207$

(b) 30/48 = 0.625

(c) The requested probability can be written in terms of whether or not the first and second washer selected are thicker than the target. That is,

$$\left(\frac{30}{50}\right)\left(\frac{29}{49}\right)\left(\frac{28}{48}\right)+\left(\frac{30}{50}\right)\left(\frac{20}{49}\right)\left(\frac{29}{48}\right)+\left(\frac{20}{50}\right)\left(\frac{30}{49}\right)\left(\frac{29}{48}\right)+\left(\frac{20}{50}\right)\left(\frac{19}{49}\right)\left(\frac{30}{48}\right)=0.60$$

2-218. Washers are selected from the lot at random without replacement.
(a) What is the minimum number of washers that need to be selected so that the probability that all the washers are thinner than the target is less than 0.10?
(b) What is the minimum number of washers that need to be selected so that the probability that 1 or more washers are thicker than the target is at least 0.90?

(a) If n washers are selected, then the probability they are all less than the target is $\dfrac{20}{50}\cdot\dfrac{19}{49}\cdot\ldots\dfrac{20-n+1}{50-n+1}$.

n	probability all selected washers are less than target
1	20/50 = 0.4
2	(20/50)(19/49) = 0.155
3	(20/50)(19/49)(18/48) = 0.058

Therefore, the answer is n = 3.

(b) Then event E that one or more washers is thicker than target is the complement of the event that all are less than target. Therefore, P(E) equals one minus the probability in part a. Therefore, n = 3.

2-219. The following table lists the history of 940 orders for features in an entry-level computer product.

		Extra Memory	
		No	Yes
Optional high-speed processor	No	514	68
	Yes	112	246

Let A be the event that an order requests the optional highspeed processor, and let B be the event that an order requests extra memory. Determine the following probabilities:

(a) $P(A \cup B)$ (b) $P(A \cap B)$ (c) $P(A' \cup B)$ (d) $P(A' \cap B')$

(e) What is the probability that an order requests an optional high-speed processor given that the order requests extra memory?

(f) What is the probability that an order requests extra memory given that the order requests an optional highspeed processor?

$$a)\quad P(A \cup B) = \frac{112 + 68 + 246}{940} = 0.453$$

$$b)\quad P(A \cap B) = \frac{246}{940} = 0.262$$

$$c)\quad P(A' \cup B) = \frac{514 + 68 + 246}{940} = 0.881$$

$$d)\quad P(A' \cap B') = \frac{514}{940} = 0.547$$

. e) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{246/940}{314/940} = 0.783$

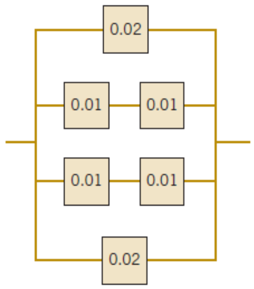
f) $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{246/940}{358/940} = 0.687$

2-220. The alignment between the magnetic media and head in a magnetic storage system affects the system's performance. Suppose that 10% of the read operations are degraded by skewed alignments, 5% of the read operations are degraded by off-center alignments, and the remaining read operations are properly aligned. The probability of a read error is 0.01 from a skewed alignment, 0.02 from an off-center alignment, and 0.001 from a proper alignment.
(a) What is the probability of a read error?
(b) If a read error occurs, what is the probability that it is due to a skewed alignment?

Let E denote a read error and let S,O,P denote skewed, off-center, and proper alignments, respectively.
Then,
(a) P(E) = P(E|S) P(S) + P(E|O) P (O) + P(E|P) P(P)
= 0.01(0.10) + 0.02(0.05) + 0.001(0.85)
= 0.00285

(b) $P(S|E) = \frac{P(E|S)P(S)}{P(E)} = \frac{0.01(0.10)}{0.00285} = 0.351$

2-221. The following circuit operates if and only if there is a path of functional devices from left to right. Assume that devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit does not operate?



Let A_i denote the event that the *i*th row operates. Then,

$$P(A_1) = 0.98, P(A_2) = (0.99)(0.99) = 0.9801, P(A_3) = 0.9801, \quad P(A_4) = 0.98.$$

The probability the circuit does not operate is

$$P(A_1')P(A_2')P(A_3')P(A_4') = (0.02)(0.0199)(0.0199)(0.02) = 1.58 \times 10^{-7}$$

2-222. A company that tracks the use of its Web site determined that the more pages a visitor views, the more likely the visitor is to provide contact information. Use the following tables to answer the questions:

Number of pages viewed:	1	2	3	4 or more
Percentage of visitors:	40	30	20	10
Percentage of visitors in each page-view category that provides contact information:	10	10	20	40

(a) What is the probability that a visitor to the Web site provides contact information?
(b) If a visitor provides contact information, what is the probability that the visitor viewed four or more pages?

(a) (0.4)(0.1) + (0.3)(0.1) +(0.2)(0.2) + (0.4)(0.1) = 0.15
(b) P(4 or more | provided info) = (0.4)(0.1)/0.15 = 0.267

2-223. An article in *Genome Research* ["An Assessment of Gene Prediction Accuracy in Large DNA Sequences" (2000, Vol. 10, pp. 1631–1642)], considered the accuracy of commercial software to predict nucleotides in gene sequences. The following table shows the number of sequences for which the programs produced predictions and the number of nucleotides correctly predicted (computed globally from the total number of prediction successes and failures on all sequences).

	Number of Sequences	Proportion
GenScan	177	0.93
Blastx default	175	0.91
Blastx topcomboN	174	0.97
Blastx 2 stages	175	0.90
GeneWise	177	0.98
Procrustes	177	0.93

Assume the prediction successes and failures are independent among the programs.
(a) What is the probability that all programs predict a nucleotide correctly?
(b) What is the probability that all programs predict a nucleotide incorrectly?
(c) What is the probability that at least one Blastx program predicts a nucleotide correctly?

(a) P=(0.93)(0.91)(0.97)(0.90)(0.98)(0.93)=0.67336
(b) P=(1-0.93)(1-0.91)(1-0.97)(1-0.90)(1-0.98)(1-0.93)=2.646×10^{-8}
(c) P=1-(1-0.91)(1-0.97)(1-0.90)=0.99973

2-224. A batch contains 36 bacteria cells. Assume that 12 of the cells are not capable of cellular replication. Of the cells, 6 are selected at random, without replacement, to be checked for replication.
(a) What is the probability that all 6 of the selected cells are able to replicate?
(b) What is the probability that at least 1 of the selected cells is not capable of replication?

(a) P=(24/36)(23/35)(22/34)(21/33)(20/32)(19/31)=0.069
(b) P=1-0.069=0.931

2-225. A computer system uses passwords that are exactly seven characters, and each character is one of the 26 letters (a–z) or 10 integers (0–9). Uppercase letters are not used.
(a) How many passwords are possible?
(b) If a password consists of exactly 6 letters and 1 number, how many passwords are possible?
(c) If a password consists of 5 letters followed by 2 numbers, how many passwords are possible?

(a) 36^7
(b) Number of permutations of six letters is 26^6. Number of ways to select one number = 10. Number of positions among the six letters to place the one number = 7. Number of passwords = $26^6 \times 10 \times 7$
(c) $26^5 10^2$

2-226. Natural red hair consists of two genes. People with red hair have two dominant genes, two regressive genes, or one dominant and one regressive gene. A group of 1000 people was categorized as follows:

	Gene 2		
Gene 1	Dominant	Regressive	Other
Dominant	5	25	30
Regressive	7	63	35
Other	20	15	800

Let *A* denote the event that a person has a dominant red hair gene, and let *B* denote the event that a person has a regressive red hair gene. If a person is selected at random from this group, compute the following:
(a) $P(A)$ (b) $P(A \cap B)$ (c) $P(A \cup B)$ (d) $P(A' \cap B)$ (e) $P(A \mid B)$
(f) Probability that the selected person has red hair

(a) $P(A) = \dfrac{5+25+30+7+20}{1000} = 0.087$

(b) $P(A \cap B) = \dfrac{25+7}{1000} = 0.032$

(c) $P(A \cup B) = 1 - \dfrac{800}{1000} = 0.20$

(d) $P(A' \cap B) = \dfrac{63+35+15}{1000} = 0.113$

(e) $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.032}{(25+63+15+7+35)/1000} = 0.2207$

(f) $P = (5+25+7+63)/1000 = 0.1$

2-227. Two suppliers each supplied 2000 parts that were evaluated for conformance to specifications. One part type was more complex than the other. The proportion of nonconforming parts of each type are shown in the table.

Supplier		Simple Component	Complex Assembly	Total
1	Nonconforming	2	10	12
	Total	1000	1000	2000
2	Nonconforming	4	6	10
	Total	1600	400	2000

One part is selected at random from each supplier. For each supplier, separately calculate the following probabilities:
(a) What is the probability a part conforms to specifications?
(b) What is the probability a part conforms to specifications given it is a complex assembly?
(c) What is the probability a part conforms to specifications given it is a simple component?
(d) Compare your answers for each supplier in part (a) to those in parts (b) and (c) and explain any unusual results.

(a) Let A denote that a part conforms to specifications and let B denote a simple component.
For supplier 1: P(A) = 1988/2000 = 0.994
For supplier 2: P(A)= 1990/2000 = 0.995
(b)
For supplier 1: P(A|B') = 990/1000 = 0.99
For supplier 2: P(A|B') = 394/400 = 0.985
(c)
For supplier 1: P(A|B) = 998/1000 = 0.998
For supplier 2: P(A|B) = 1596/1600 = 0.9975

(d) The unusual result is that for both a simple component and for a complex assembly, supplier 1 has a greater probability that a part conforms to specifications. However, supplier 1 has a lower probability of conformance overall. The overall conforming probability depends on both the conforming probability of each part type and also the probability of each part type. Supplier 1 produces more of the complex parts so that overall conformance from supplier 1 is lower.

2-228. The article "Term Efficacy of Ribavirin Plus Interferon Alfa in the Treatment of Chronic Hepatitis C," [*Gastroenterology* (1996, Vol. 111, no. 5, pp. 1307–1312)], considered the effect of two treatments and a control for treatment of hepatitis C. The following table provides the total patients in each group and the number that showed a complete (positive) response after 24 weeks of treatment. Suppose a patient is selected randomly.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let *A* denote the event that the patient is treated with ribavirin plus interferon alfa or interferon alfa, and let *B* denote the event that the response is complete. Determine the following probabilities.

(a) $P(A \mid B)$ (b) $P(B \mid A)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

P(A|B)= $\frac{P(B|A)P(A)}{P(B)} = \frac{\frac{22}{40}(\frac{40}{60})}{\frac{22}{60}} = 1$
P(B|A) = 22/40 = 0.55
P(A∩B) = P(B|A)P(A)=(22/40)(40/60) = 22/60 = 0.366
P(AUB) = P(A)+P(B) - P(A∩B) = (40/60) + (22/60) – (22/60) = 0.667

2-229. The article ["Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis," *Arthritis & Rheumatism* (2005, Vol. 52, pp. 3381– 3390)] considered four treatment groups. The groups consisted of patients with different drug therapies (such as prednisone and infliximab): sequential monotherapy (group 1), step-up combination therapy (group 2), initial combination therapy (group 3), or initial combination therapy with infliximab (group 4). Radiographs of hands and feet were used to evaluate disease progression. The number of patients without progression of joint damage was 76 of 114 patients (67%), 82 of 112 patients (73%), 104 of 120 patients (87%), and 113 of 121 patients (93%) in groups 1–4, respectively. Suppose a patient is selected randomly. Let *A* denote the event that the patient is in group 1 or 2, and let *B* denote the event that there is no progression. Determine the following probabilities:

(a) $P(A \mid B)$ (b) $P(B \mid A)$ (c) $P(A \cap B)$ (d) $P(A \cup B)$

(a) P(A|B)= $\frac{158}{375} = 0.421$
(b) P(B|A)= $\frac{158}{226} = 0.699$
(c) P(A∩B)= $\frac{76+82}{467} = 0.338$
(d) P(AUB)= $\frac{226}{467} + \frac{375}{467} - \frac{158}{467}$ = 0.948

Mind-Expanding Exercises

2-230. Suppose documents in a lending organization are selected randomly (without replacement) for review. In a set of 50 documents, suppose that 2 actually contain errors.
(a) What is the minimum sample size such that the probability exceeds 0.90 that at least 1 document in error is selected?
(b) Comment on the effectiveness of sampling inspection to detect errors.

(a) Let X denote the number of documents in error in the sample and let n denote the sample size.

$P(X \geq 1) = 1 - P(X = 0)$ and $P(X = 0) = \dfrac{\binom{2}{0}\binom{48}{n}}{\binom{50}{n}}$

Trials for n result in the following results

n	$P(X = 0)$	$1 - P(X = 0)$
5	0.808163265	0.191836735
10	0.636734694	0.363265306
15	0.485714286	0.514285714
20	0.355102041	0.644897959
25	0.244897959	0.755102041
30	0.155102041	0.844897959
33	0.111020408	0.888979592
34	0.097959184	0.902040816

Therefore $n = 34$.

(b) A large proportion of the set of documents needs to be inspected in order for the probability of a document in error to be detected to exceed 0.9.

2-231. Suppose that a lot of washers is large enough that it can be assumed that the sampling is done with replacement. Assume that 60% of the washers exceed the target thickness.
(a) What is the minimum number of washers that need to be selected so that the probability that none is thicker than the target is less than 0.10?
(b) What is the minimum number of washers that need to be selected so that the probability that 1 or more washers are thicker than the target is at least 0.90?

Let n denote the number of washers selected.
(a) The probability that none are thicker, that is, all are less than the target is 0.4^n by independence.
The following results are obtained:

n	0.4^n
1	0.4
2	0.16
3	0.064

Therefore, $n = 3$.

(b) The requested probability is the complement of the probability requested in part a). Therefore, $n = 3$

2-232. A biotechnology manufacturing firm can produce diagnostic test kits at a cost of \$20. Each kit for which there is a demand in the week of production can be sold for \$100. However, the half-life of components in the kit requires the kit to be scrapped if it is not sold in the week of production. The cost of scrapping the kit is \$5. The weekly demand is summarized as follows:

Weekly Demand				
Number of units	0	50	100	200
Probability of demand	0.05	0.4	0.3	0.25

How many kits should be produced each week to maximize the firm's mean earnings?

Let x denote the number of kits produced.

Revenue at each demand

	0	50	100	200
$0 \le x \le 50$	-5x	100x	100x	100x
Mean profit = 100x(0.95)-5x(0.05)-20x				
$50 \le x \le 100$	-5x	100(50)-5(x-50)	100x	100x
Mean profit = [100(50)-5(x-50)](0.4) + 100x(0.55)-5x(0.05)-20x				
$100 \le x \le 200$	-5x	100(50)-5(x-50)	100(100)-5(x-100)	100x
Mean profit = [100(50)-5(x-50)](0.4) + [100(100)-5(x-100)](0.3) + 100x(0.25) - 5x(0.05) - 20x				

	Mean Profit	Maximum Profit
$0 \le x \le 50$	74.75 x	$ 3737.50 at x=50
$50 \le x \le 100$	32.75 x + 2100	$ 5375 at x=100
$100 \le x \le 200$	1.25 x + 5250	$ 5500 at x=200

Therefore, profit is maximized at 200 kits. However, the difference in profit over 100 kits is small.

2-233. A steel plate contains 20 bolts. Assume that 5 bolts are not torqued to the proper limit. 4 bolts are selected at random, without replacement, to be checked for torque. If an operator checks a bolt,the probability that an incorrectly torqued bolt is identified is 0.95. If a checked bolt is correctly torqued, the operator's conclusion is always correct. What is the probability that at least one bolt in the sample of four is identified as being incorrectly torqued?

Let E denote the event that none of the bolts are identified as incorrectly torqued.
Let X denote the number of bolts in the sample that are incorrect. The requested probability is P(E').
Then,
P(E) = P(E|X=0)P(X=0) + P(E|X=1) P(X=1) + P(E|X=2) P(X=2) + P(E|X=3) P(X=3) + P(E|X=4)P(X=4)
and P(X=0) = (15/20)(14/19)(13/18)(12/17) = 0.2817.
The remaining probability for X can be determined from the counting methods. Then

$$P(X=1)=\frac{\binom{5}{1}\binom{15}{3}}{\binom{20}{4}}=\frac{\left(\frac{5!}{4!1!}\right)\left(\frac{15!}{3!12!}\right)}{\left(\frac{20!}{4!16!}\right)}=\frac{5!15!4!16!}{4!3!12!20!}=0.4696$$

$$P(X=2)=\frac{\binom{5}{2}\binom{15}{2}}{\binom{20}{4}}=\frac{\left(\frac{5!}{3!2!}\right)\left(\frac{15!}{2!13!}\right)}{\left(\frac{20!}{4!16!}\right)}=0.2167$$

$$P(X=3)=\frac{\binom{5}{3}\binom{15}{1}}{\binom{20}{4}}=\frac{\left(\frac{5!}{3!2!}\right)\left(\frac{15!}{1!14!}\right)}{\left(\frac{20!}{4!16!}\right)}=0.0309$$

P(X = 4) = (5/20)(4/19)(3/18)(2/17) = 0.0010, P(E | X = 0) = 1, P(E | X = 1) = 0.05,
P(E | X = 2) = 0.05^2 = 0.0025, P(E|X=3) = 0.05^3 = $1.25\text{x}10^{-4}$, P(E | X=4) = 0.05^4 = $6.25\text{x}10^{-6}$.
Then, $P(E)=1(0.2817)+0.05(0.4696)+0.0025(0.2167)+1.25\times10^{-4}(0.0309)+6.25\times10^{-6}(0.0010)=0.306$
and P(E') = 0.694

2-234. If the events *A* and *B* are independent, show that *A′* and *B′* are independent.

$$\begin{aligned}P(A'\cap B') &= 1-P([A'\cap B']')=1-P(A\cup B)\\ &=1-[P(A)+P(B)-P(A\cap B)]\\ &=1-P(A)-P(B)+P(A)P(B)\\ &=[1-P(A)][1-P(B)]\\ &=P(A')P(B')\end{aligned}$$

2-235. Suppose that a table of part counts is generalized as follows:

		Conforms	
		Yes	No
Supplier	1	*ka*	*kb*
	2	*a*	*b*

where a, b, and k are positive integers. Let A denote the event that a part is from supplier 1, and let B denote the event that a part conforms to specifications. Show that A and B are independent events. This exercise illustrates the result that whenever the rows of a table (with r rows and c columns) are proportional, an event defined by a row category and an event defined by a column category are independent.

The total sample size is $ka + a + kb + b = (k + 1)a + (k + 1)b$. Therefore

$$P(A) = \frac{k(a+b)}{(k+1)a+(k+1)b}, P(B) = \frac{ka+a}{(k+1)a+(k+1)b}$$

and

$$P(A \cap B) = \frac{ka}{(k+1)a+(k+1)b} = \frac{ka}{(k+1)(a+b)}$$

Then,

$$P(A)P(B) = \frac{k(a+b)(ka+a)}{[(k+1)a+(k+1)b]^2} = \frac{k(a+b)(k+1)a}{(k+1)^2(a+b)^2} = \frac{ka}{(k+1)(a+b)} = P(A \cap B)$$

CHAPTER 3

Section 3-1

3-1. The range of X is $\{0,1,2,...,1000\}$

3-2. The range of X is $\{0,1,2,\ldots,50\}$

3-3. The range of X is $\{0,1,2,\ldots,99999\}$

3-4. The range of X is {0, 1, 2, 3, 4, 5}

3-5. The range of X is {1, 2, …, 491}. Because 490 parts are conforming, a nonconforming part must be selected in 491 selections.

3-6. The range of X is $\{0,1,2,\ldots,100\}$. Although the range actually obtained from lots typically might not exceed 10%.

3-7. The range of X is conveniently modeled as all nonnegative integers. That is, the range of X is {0, 1, 2, …}

3-8. The range of X is conveniently modeled as all nonnegative integers. That is, the range of X is {0, 1, 2, …}

3-9. The range of X is {0, 1, 2, …, 15}

3-10. The possible totals for two orders are 1/8 + 1/8 = 1/4, 1/8 + 1/4 = 3/8, 1/8 + 3/8 = 1/2, 1/4 + 1/4 = 1/2, 1/4 + 3/8 = 5/8, 3/8 + 3/8 = 6/8.
Therefore the range of X is $\left\{\frac{1}{4},\frac{3}{8},\frac{1}{2},\frac{5}{8},\frac{6}{8}\right\}$

3-11. The range of X is {0, 1, 2, …, 10,000}

3-12. The range of X is {100, 101, …, 150}

3-13. The range of X is {0, 1, 2, …, 40000)

3-14. The range of X is {0, 1, 2, ..., 16}.

3-15. The range of X is {1, 2, ..., 100}.

Section 3-2

3-16.

$$f_X(0) = P(X = 0) = 1/6 + 1/6 = 1/3$$
$$f_X(1.5) = P(X = 1.5) = 1/3$$
$$f_X(2) = 1/6$$
$$f_X(3) = 1/6$$

a) P(X = 1.5) = 1/3
b) P(0.5< X < 2.7) = P(X = 1.5) +P(X = 2) = 1/3 + 1/6 = 1/2
c) P(X > 3) = 0
d) $P(0 \le X < 2) = P(X = 0) + P(X = 1.5) = 1/3 + 1/3 = 2/3$
e) P(X = 0 or X = 2) = 1/3 + 1/6 = 1/2

3-17. All probabilities are greater than or equal to zero and sum to one.
a) P(X ≤ 2)=1/8 + 2/8 + 2/8 + 2/8 + 1/8 = 1
b) P(X > - 2) = 2/8 + 2/8 + 2/8 + 1/8 = 7/8
c) P(-1 ≤ X ≤ 1) = 2/8 + 2/8 + 2/8 =6/8 = 3/4

d) $P(X \le -1 \text{ or } X=2) = 1/8 + 2/8 + 1/8 = 4/8 = 1/2$

3-18. All probabilities are greater than or equal to zero and sum to one.
a) $P(X\le 1)=P(X=1)=0.5714$
b) $P(X>1)= 1-P(X=1)=1-0.5714=0.4286$
c) $P(2<X<6)=P(X=3)=0.1429$
d) $P(X\le 1 \text{ or } X>1)= P(X=1)+ P(X=2)+P(X=3)=1$

3-19. Probabilities are nonnegative and sum to one.
a) $P(X = 4) = 9/25$
b) $P(X \le 1) = 1/25 + 3/25 = 4/25$
c) $P(2 \le X < 4) = 5/25 + 7/25 = 12/25$
d) $P(X > -10) = 1$

3-20. Probabilities are nonnegative and sum to one.
a) $P(X = 2) = 3/4(1/4)^2 = 3/64$
b) $P(X \le 2) = 3/4[1+1/4+(1/4)^2] = 63/64$
c) $P(X > 2) = 1 - P(X \le 2) = 1/64$
d) $P(X \ge 1) = 1 - P(X \le 0) = 1 - (3/4) = 1/4$

3-21.
a) $P(X\ge 2)=P(X=2)+P(X=2.25)=0.2+0.1=0.3$
b) $P(X<1.65)=P(X=1.25)+P(X=1.5)=0.2+0.4=0.6$
c) $P(X=1.5)=f(1.5)=0.4$
d) $P(X<1.3 \text{ or } X>2.1)=P(X=1.25)+P(X=2.25)=0.2+0.1=0.3$

3-22. X = the number of patients in the sample who are admitted
Range of X = {0,1,2}
A = the event that the first patient is admitted
B = the event that the second patient is admitted
A and B are independent events due to the selection with replacement.

$P(A)=P(B)=1277/5292=0.2413$
$P(X=0) = P(A'\cap B') = (1-0.2413)(1-0.2413) = 0.576$
$P(X=1) = P(A \cap B')+P(A'\cap B) = 0.2413(1-0.2413)+ (1-0.2413)(0.2413) = 0.366$
$P(X=2) = (A\cap B) = 0.2413\times 0.2413 = 0.058$

x	P(X=x)
0	0.576
1	0.366
2	0.058

3-23. X = number of successful surgeries.
$P(X=0)=0.1(0.33)=0.033$
$P(X=1)=0.9(0.33)+0.1(0.67)=0.364$
$P(X=2)=0.9(0.67)=0.603$

3-24. $P(X = 0) = 0.02^3 = 8 \times 10^{-6}$
$P(X = 1) = 3[0.98(0.02)(0.02)]=0.0012$
$P(X = 2) = 3[0.98(0.98)(0.02)]=0.0576$
$P(X = 3) = 0.98^3 = 0.9412$

3-25. X = number of wafers that pass
$P(X=0) = (0.2)^3 = 0.008$
$P(X=1) = 3(0.2)^2(0.8) = 0.096$
$P(X=2) = 3(0.2)(0.8)^2 = 0.384$
$P(X=3) = (0.8)^3 = 0.512$

3-26. X: the number of computers that vote for a left roll when a right roll is appropriate.
p=0.0001.
$P(X=0)=(1-p)^4=0.9999^4=0.9996$

$P(X=1)=4*(1-p)^3p=4*0.9999^3 0.0001=0.0003999$
$P(X=2)=C_4^2(1-p)^2p^2=5.999*10^{-8}$
$P(X=3)=C_4^3(1-p)^1p^3=3.9996*10^{-12}$
$P(X=4)=C_4^0(1-p)^0p^4=1*10^{-16}$

3-27. P(X = 50 million) = 0.5, P(X = 25 million) = 0.3, P(X = 10 million) = 0.2

3-28. P(X = 10 million) = 0.3, P(X = 5 million) = 0.6, P(X = 1 million) = 0.1

3-29. P(X = 15 million) = 0.6, P(X = 5 million) = 0.3, P(X = -0.5 million) = 0.1

3-30. X = number of components that meet specifications
P(X=0) = (0.05)(0.02) = 0.001
P(X=1) = (0.05)(0.98) + (0.95)(0.02) = 0.068
P(X=2) = (0.95)(0.98) = 0.931

3-31. X = number of components that meet specifications
P(X=0) = (0.05)(0.02)(0.01) = 0.00001
P(X=1) = (0.95)(0.02)(0.01) + (0.05)(0.98)(0.01)+(0.05)(0.02)(0.99) = 0.00167
P(X=2) = (0.95)(0.98)(0.01) + (0.95)(0.02)(0.99) + (0.05)(0.98)(0.99) = 0.07663
P(X=3) = (0.95)(0.98)(0.99) = 0.92169

3-32. X = final temperature
P(X=266) = 48/200 = 0.24
P(X=271) = 60/200 = 0.30
P(X=274) = 92/200 = 0.46

$$f(x) = \begin{cases} 0.24, & x = 266 \\ 0.30, & x = 271 \\ 0.46, & x = 274 \end{cases}$$

3-33. X = waiting time (hours)
P(X=1) = 19/500 = 0.038
P(X=2) = 51/500 = 0.102
P(X=3) = 86/500 = 0.172
P(X=4) = 102/500 = 0.204
P(X=5) = 87/500 = 0.174
P(X=6) = 62/500 = 0.124
P(X=7) = 40/500 = 0.08
P(X=8) = 18/500 = 0.036
P(X=9) = 14/500 = 0.028
P(X=10) = 11/500 = 0.022
P(X=15) = 10/500 = 0.020

$$f(x)=\begin{cases}0.038, & x=1\\ 0.102, & x=2\\ 0.172, & x=3\\ 0.204, & x=4\\ 0.174, & x=5\\ 0.124, & x=6\\ 0.080, & x=7\\ 0.036, & x=8\\ 0.028, & x=9\\ 0.022, & x=10\\ 0.020, & x=15\end{cases}$$

3-34. X = days until change
P(X=1.5) = 0.05
P(X=3) = 0.25
P(X=4.5) = 0.35
P(X=5) = 0.20
P(X=7) = 0.15

$$f(x)=\begin{cases}0.05, & x=1.5\\ 0.25, & x=3\\ 0.35, & x=4.5\\ 0.20, & x=5\\ 0.15, & x=7\end{cases}$$

3-35. X = Non-failed well depth
P(X=255) = (1515+1343)/7726 = 0.370
P(X=218) = 26/7726 = 0.003
P(X=317) = 3290/7726 = 0.426
P(X=231) = 349/7726 = 0.045
P(X=267) = (280+887)/7726 = 0.151
P(X=217) = 36/7726 = 0.005

$$f(x)=\begin{cases}0.005, & x=217\\ 0.003, & x=218\\ 0.045, & x=231\\ 0.370, & x=255\\ 0.151, & x=267\\ 0.426, & x=317\end{cases}$$

3-36. X = the number of wafers selected
X ϵ {1, 2, ...}
P(X=1) = 0.10
P(X=2) = (1−0.10)(0.10) = 0.09
P(X=3) = $(1-0.10)^2(0.10)$ = 0.081
P(X=4) = $(1-0.10)^3(0.10)$ = 0.0729
In general, P(X=x) = $(0.10)\,(1-0.10)^{x-1}$ for x = 1, 2, 3,…

3-37. X = the number of failed devices
Range of X = {0,1,2}

A = the event that the first device fails
B = the event that the second device fails
$P(X=0) = P(A'\cap B') = (0.8)(0.9) = 0.72$
$P(X=1) = P(A\cap B') + P(A'\cap B) = (1-0.8)(0.9)+(0.8)(1-0.9) = 0.26$
$P(X=2) = P(A\cap B) = (1-0.8)(1-0.9) = 0.02$

x	P(X=x)
0	0.72
1	0.26
2	0.02

Section 3-3

3-38. $$F(x)=\begin{cases} 0, & x<0 \\ 1/3 & 0\le x<1.5 \\ 2/3 & 1.5\le x<2 \\ 5/6 & 2\le x<3 \\ 1 & 3\le x \end{cases}$$ where
$$f_X(0)=P(X=0)=1/6+1/6=1/3$$
$$f_X(1.5)=P(X=1.5)=1/3$$
$$f_X(2)=1/6$$
$$f_X(3)=1/6$$

3-39.
$$F(x)=\begin{cases} 0, & x<-2 \\ 1/8 & -2\le x<-1 \\ 3/8 & -1\le x<0 \\ 5/8 & 0\le x<1 \\ 7/8 & 1\le x<2 \\ 1 & 2\le x \end{cases}$$ where
$$f_X(-2)=1/8$$
$$f_X(-1)=2/8$$
$$f_X(0)=2/8$$
$$f_X(1)=2/8$$
$$f_X(2)=1/8$$

a) $P(X \le 1.25) = 7/8$
b) $P(X \le 2.2) = 1$
c) $P(-1.1 < X \le 1) = 7/8 - 1/8 = 3/4$
d) $P(X > 0) = 1 - P(X \le 0) = 1 - 5/8 = 3/8$

3-40.
$$F(x)=\begin{cases} 0 & x<1 \\ 4/7 & 1\le x<2 \\ 6/7 & 2\le x<3 \\ 1 & 3\le x \end{cases}$$

a) $P(X < 1.5) = 4/7$
b) $P(X \le 3) = 1$
c) $P(X > 2) = 1 - P(X \le 2) = 1 - 6/7 = 1/7$
d) $P(1 < X \le 2) = P(X \le 2) - P(X \le 1) = 6/7 - 4/7 = 2/7$

3-41.

$$F(x)=\begin{cases} 0, & x<0 \\ 1/25 & 0\le x<1 \\ 4/25 & 1\le x<2 \\ 9/25 & 2\le x<3 \\ 16/25 & 3\le x<4 \\ 1 & 4\le x \end{cases}$$

3-42.

$$F(x)=\begin{cases} 0, & x<0 \\ 3/4 & 0\le x<1 \\ 15/16 & 1\le x<2 \\ 63/64 & 2\le x<3 \\ \dots & \dots \\ 1 & x<\infty \end{cases}$$

In general, $F(x) = (4^k-1) / 4^k$ for $k - 1 \le x < k$

3-43.

$$F(x)=\begin{cases} 0 & x<1.25 \\ 0.2 & 1.25\le x<1.5 \\ 0.6 & 1.5\le x<1.75 \\ 0.7 & 1.75\le x<2 \\ 0.9 & 2\le x<2.25 \\ 1 & 2.25\le x \end{cases}$$

3-44.

Probability a patient from hospital 1 is admitted is 1277/5292 = 0.2413
Here X is the number of patents admitted in the sample.
Range of X = {0,1,2}
P(X=0) = (1 − 0.2413)(1 − 0.2413) = 0.576
P(X=1) = 0.2413(1 − 0.2413)+ (1 − 0.2413)(0.2413) = 0.366
P(X=2) = 0.2413×0.2413 = 0.058
The cumulative distribution function of X is

$$F(x)=\begin{cases} 0 & x<0 \\ 0.576 & 0\le x<1 \\ 0.942 & 1\le x<2 \\ 1 & x\le 2 \end{cases}$$

3-45.

$$F(x)=\begin{cases} 0, & x<0 \\ 0.008, & 0\le x<1 \\ 0.104, & 1\le x<2 \\ 0.488, & 2\le x<3 \\ 1, & 3\le x \end{cases}$$

$$f(0) = 0.2^3 = 0.008$$
$$f(1) = 3(0.2)(0.2)(0.8) = 0.096$$
$$f(2) = 3(0.2)(0.8)(0.8) = 0.384$$
$$f(3) = (0.8)^3 = 0.512$$

3-46.

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.9996, & 0 \le x < 1 \\ 0.9999, & 1 \le x < 3 \\ 0.99999, & 3 \le x < 4 \\ 1, & 4 \le x \end{cases}$$

$$f(0) = 0.9999^4 = 0.9996$$
$$f(1) = 4(0.9999^3)(0.0001) = 0.000399$$
$$f(2) = 5.999(10^{-8})$$
$$f(3) = 3.9996(10^{-12})$$
$$f(4) = 10^{-16}$$

3-47.

$$F(x) = \begin{cases} 0, & x < 10 \\ 0.2, & 10 \le x < 25 \\ 0.5, & 25 \le x < 50 \\ 1, & 50 \le x \end{cases}$$

where P(X = 50 million) = 0.5, P(X = 25 million) = 0.3, P(X = 10 million) = 0.2

3-48.

$$F(x) = \begin{cases} 0, & x < 1 \\ 0.1, & 1 \le x < 5 \\ 0.7, & 5 \le x < 10 \\ 1, & 10 \le x \end{cases}$$

where P(X = 10 million) = 0.3, P(X = 5 million) = 0.6, P(X = 1 million) = 0.1

3-49. The sum of the probabilities is 1 and all probabilities are greater than or equal to zero;
f(1) = 0.5, *f*(3) = 0.5
a) $P(X \le 3) = 1$
b) $P(X \le 2) = 0.5$
c) $P(1 \le X \le 2) = P(X=1) = 0.5$
d) $P(X>2) = 1 - P(X\le 2) = 0.5$

3-50. The sum of the probabilities is 1 and all probabilities are greater than or equal to zero;
f(1) = 0.7, f(4) = 0.2, f(7) = 0.1
a) $P(X \le 4) = 0.9$
b) $P(X > 7) = 0$
c) $P(X \le 5) = 0.9$

d) P(X>4) = 0.1
e) P(X≤2) = 0.7

3-51. The sum of the probabilities is 1 and all probabilities are greater than or equal to zero;
f(-10) = 0.25, f(30) = 0.5, f(50) = 0.25
a) P(X≤50) = 1
b) P(X≤40) = 0.75
c) P(40 ≤ X ≤ 60) = P(X=50)=0.25
d) P(X<0) = 0.25
e) P(0≤X<10) = 0
f) P(−10<X<10) = 0

3-52. The sum of the probabilities is 1 and all probabilities are greater than or equal to zero;
f(1/8) = 0.2, f(1/4) = 0.7, f(3/8) = 0.1
a) P(X≤1/18) = 0
b) P(X≤1/4) = 0.9
c) P(X≤5/16) = 0.9
d) P(X>1/4) = 0.1
e) P(X≤1/2) = 1

3-53.

$$F(x) = \begin{cases} 0, & x < 266 \\ 0.24, & 266 \le x < 271 \\ 0.54, & 271 \le x < 274 \\ 1, & 274 \le x \end{cases}$$

where P(X = 266 K) = 0.24, P(X = 271 K) = 0.30, P(X = 274 K) = 0.46

3-54.

$$F(x) = \begin{cases} 0, & x < 1 \\ 0.038, & 1 \le x < 2 \\ 0.140, & 2 \le x < 3 \\ 0.312, & 3 \le x < 4 \\ 0.516, & 4 \le x < 5 \\ 0.690, & 5 \le x < 6 \\ 0.814, & 6 \le x < 7 \\ 0.894, & 7 \le x < 8 \\ 0.930, & 8 \le x < 9 \\ 0.958, & 9 \le x < 10 \\ 0.980, & 10 \le x < 15 \\ 1 & 15 \le x \end{cases}$$

where P(X=1) = 0.038, P(X=2) = 0.102, P(X=3) = 0.172, P(X=4) = 0.204, P(X=5) = 0.174, P(X=6) = 0.124, P(X=7) = 0.08, P(X=8) = 0.036, P(X=9) = 0.028, P(X=10) = 0.022, P(X=15) = 0.020

3-55.

$$F(x) = \begin{cases} 0, & x < 1.5 \\ 0.05, & 1.5 \le x < 3 \\ 0.30, & 3 \le x < 4.5 \\ 0.65, & 4.5 \le x < 5 \\ 0.85, & 5 \le x < 7 \\ 1 & 7 \le x \end{cases}$$

where P(X=1.5) = 0.05, P(X= 3) = 0.25, P(X=4.5) = 0.35, P(X=5) = 0.20, P(X=7) = 0.15

3-56.

$$F(x) = \begin{cases} 0, & x < 217 \\ 0.005, & 217 \le x < 218 \\ 0.008, & 218 \le x < 231 \\ 0.053, & 231 \le x < 255 \\ 0.423, & 255 \le x < 267 \\ 0.574, & 267 \le x < 317 \\ 1, & 317 \le x \end{cases}$$

where P(X=255) = 0.370, P(X=218) = 0.003, P(X=317) = 0.426, P(X=231) = 0.045, P(X=267) = 0.151, P(X=217) = 0.005

<u>Section 3-4</u>

3-57. Mean and Variance

$$\mu = E(X) = 0f(0) + 1f(1) + 2f(2) + 3f(3) + 4f(4)$$
$$= 0(0.2) + 1(0.2) + 2(0.2) + 3(0.2) + 4(0.2) = 2$$
$$V(X) = 0^2 f(0) + 1^2 f(1) + 2^2 f(2) + 3^2 f(3) + 4^2 f(4) - \mu^2$$
$$= 0(0.2) + 1(0.2) + 4(0.2) + 9(0.2) + 16(0.2) - 2^2 = 2$$

3- 58. Mean and Variance for random variable in exercise 3-14

$$\mu = E(X) = 0f(0) + 1.5f(1.5) + 2f(2) + 3f(3)$$
$$= 0(1/3) + 1.5(1/3) + 2(1/6) + 3(1/6) = 1.333$$
$$V(X) = 0^2 f(0) + 1.5^2 f(1) + 2^2 f(2) + 3^2 f(3) - \mu^2$$
$$= 0(1/3) + 2.25(1/3) + 4(1/6) + 9(1/6) - 1.333^2 = 1.139$$

3-59. Determine E(X) and V(X) for random variable in exercise 3-15

$$\mu = E(X) = -2f(-2) - 1f(-1) + 0f(0) + 1f(1) + 2f(2)$$
$$= -2(1/8) - 1(2/8) + 0(2/8) + 1(2/8) + 2(1/8) = 0$$
$$V(X) = -2^2 f(-2) - 1^2 f(-1) + 0^2 f(0) + 1^2 f(1) + 2^2 f(2) - \mu^2$$
$$= 4(1/8) + 1(2/8) + 0(2/8) + 1(2/8) + 4(1/8) - 0^2 = 1.5$$

3-60. Determine E(X) and V(X) for random variable in exercise 3-16

$$\mu = E(X) = 1f(1) + 2f(2) + 3f(3)$$
$$= 1(0.5714286) + 2(0.2857143) + 3(0.1428571)$$
$$= 1.571429$$
$$V(X) = 1^2 f(1) + 2^2 f(2) + 3^2 f(3) + 4^2 f(4) - \mu^2 = 0.0531$$

3-61.

$$\mu = E(X) = 0f(0) + 1f(1) + 2f(2) + 3f(3) + 4f(4)$$
$$= 0(0.04) + 1(0.12) + 2(0.2) + 3(0.28) + 4(0.36) = 2.8$$
$$V(X) = 0^2 f(0) + 1^2 f(1) + 2^2 f(2) + 3^2 f(3) + 4^2 f(4) - \mu^2$$
$$= 0(0.04) + 1(0.12) + 4(0.2) + 9(0.28) + 16(0.36) - 2.8^2 = 1.36$$

3-62.

$$E(X) = \frac{3}{4}\sum_{x=0}^{\infty} x\left(\frac{1}{4}\right)^x = \frac{3}{4}\sum_{x=1}^{\infty} x\left(\frac{1}{4}\right)^x = \frac{1}{3}$$

The result uses a formula for the sum of an infinite series. The formula can be derived from the fact that the series to sum is the derivative of $h(a) = \sum_{x=1}^{\infty} a^x = \frac{a}{1-a}$ with respect to *a*.

For the variance, another formula can be derived from the second derivative of *h*(*a*) with respect to *a*. Calculate from this formula

$$E(X^2) = \frac{3}{4}\sum_{x=0}^{\infty} x^2\left(\frac{1}{4}\right)^x = \frac{3}{4}\sum_{x=1}^{\infty} x^2\left(\frac{1}{4}\right)^x = \frac{5}{9}$$

Then $V(X) = E(X^2) - [E(X)]^2 = \frac{5}{9} - \frac{1}{9} = \frac{4}{9}$

3-63.

$$\mu = E(X) = 0f(0) + 1f(1) + 2f(2)$$
$$= 0(0.033) + 1(0.364) + 2(0.603)$$
$$= 1.57$$
$$V(X) = 0^2 f(0) + 1^2 f(1) + 2^2 f(2) - \mu^2$$
$$= 0(0.033) + 1(0.364) + 4(0.603) - 1.57^2$$
$$= 0.3111$$

3-64.

$$\mu = E(X) = 0f(0) + 1f(1) + 2f(2) + 3f(3)$$
$$= 0(8\times10^{-6}) + 1(0.0012) + 2(0.0576) + 3(0.9412)$$
$$= 2.940008$$

$$V(X) = 0^2 f(0) + 1^2 f(1) + 2^2 f(2) + 3^2 f(3) - \mu^2$$
$$= 0.05876096$$

3-65. Determine x where range is [0, 1, 2, 3, x] and the mean is 6.

$$\mu = E(X) = 6 = 0f(0) + 1f(1) + 2f(2) + 3f(3) + xf(x)$$
$$6 = 0(0.2) + 1(0.2) + 2(0.2) + 3(0.2) + x(0.2)$$
$$6 = 1.2 + 0.2x$$
$$4.8 = 0.2x$$
$$x = 24$$

3-66. (a) F(0)=0.17

Nickel Charge: X	CDF
0	0.17
2	0.17+0.35=0.52
3	0.17+0.35+0.33=0.85
4	0.17+0.35+0.33+0.15=1

(b)E(X) = 0(0.17)+2(0.35)+3(0.33)+4(0.15) = 2.29

V(X) = $\sum_{i=1}^{4} f(x_i)(x_i - \mu)^2$ = 1.5259

3-67. X = number of computers that vote for a left roll when a right roll is appropriate.

μ = E(X) = 0f(0) + 1f(1) + 2f(2) + 3f(3) + 4f(4)

= 0 + 0.0003999 + 2(5.999×10^{-8}) + 3(3.9996×10^{-12}) + 4(1)10^{-16} = 0.0004

V(X)= $\sum_{i=1}^{5} f(x_i)(x_i - \mu)^2$ = 0.0004

3-68. μ = E(X) = 350(0.06) + 450(0.1) + 550(0.47) + 650(0.37) = 565

V(X)= $\sum_{i=1}^{4} f(x_i)(x - \mu)^2$ = 6875

σ= $\sqrt{V(X)}$ =82.92

3-69. (a)

Transaction	Frequency	Selects: X	f(X)
New order	43	23	0.43
Payment	44	4.2	0.44
Order status	4	11.4	0.04
Delivery	5	130	0.05
Stock level	4	0	0.04
Total	100		

E(X) = μ = 23(0.43) + 4.2(0.44) + 11.4(0.04) + 130(0.05) + 0(0.04) =18.694

V(X) = $\sum_{i=1}^{5} f(x_i)(x - \mu)^2$ = 735.964 $\sigma = \sqrt{V(X)} = 27.1287$

(b)

Transaction	Frequency	All operation: X	f(X)

New order	43	23+11+12=46	0.43
Payment	44	4.2+3+1+0.6=8.8	0.44
Order status	4	11.4+0.6=12	0.04
Delivery	5	130+120+10=260	0.05
Stock level	4	0+1=1	0.04
total	100		

μ = E(X) = 46*0.43+8.8*0.44+12*0.04+260*0.05+1*0.04=37.172

V(X) = $\sum_{i=1}^{5} f(x_i)(x-\mu)^2$ =2947.996 $\sigma = \sqrt{V(X)} = 54.2955$

3-70. μ = E(X) = 266(0.24) + 271(0.30) + 274(0.46) = 271.18

V(X) = $\sum_{i=1}^{5} f(x_i)(x-\mu)^2$ = 10.11

3-71. μ = E(X) = 1(0.038) + 2(0.102) + 3(0.172) + 4(0.204) + 5(0.174) + 6(0.124) + 7(0.08) + 8(0.036) + 9(0.028) + 10(0.022) +15(0.020)

= 4.808 hours

V(X) = $\sum_{i=1}^{5} f(x_i)(x-\mu)^2$ = 6.147

3-72. μ = E(X) = 1.5(0.05) + 3(0.25) + 4.5(0.35) + 5(0.20) + 7(0.15) = 4.45

V(X) = $\sum_{i=1}^{5} f(x_i)(x-\mu)^2$ = 1.9975

3-73.

X = the depth of a non-failed well

x	f(x)	xf(x)	$(x-\mu)^2 f(x)$
217	0.0047=36/7726	1.011131245	19.5831
218	0.0034=26/7726	0.733626715	13.71039
231	0.0452=349/7726	10.43476573	116.7045
255	0.3699=(1515+1343)/7726	94.32953663	266.2591
267	0.1510=887/7726	40.32992493	33.21378
317	0.4258=3290/7726	134.9896454	526.7684

μ = E(X) = 255(0.370) + 218(0.003) + 317(0.426) + 231(0.045) + 267(0.151) + 217(0.005) = 281.83

V(X) = $\sum_{i=1}^{5} f(x_i)(x-\mu)^2$ = 976.24

3-74. $f(x) = 0.1(0.9)^{x-1}$

Mean = $\sum_{x=1}^{\infty} 0.1(0.9)^{x-1} x = 0.1\sum_{x=1}^{\infty} x(0.9)^{x-1} = 0.1\frac{1}{(1-0.9)^2} = 10$

Note that $\sum_{x=1}^{\infty} xq^{x-1} = \frac{\partial}{\partial q}\sum_{x=1}^{\infty} q^x = \frac{\partial}{\partial q}\left(\frac{q}{1-q}\right) = \frac{1}{(1-q)^2}$ where $q = 0.9$

For the variance, consider

$$\sum_{x=1}^{\infty} x(x-1)q^{x-2} = \frac{\partial^2}{\partial q^2}\sum_{x=1}^{\infty} q^x = \frac{\partial^2}{\partial q^2}\left(\frac{q}{1-q}\right) = \frac{2}{(1-q)^3}$$

$$\sum_{x=1}^{\infty} x^2 q^{x-2} - \sum_{x=1}^{\infty} x q^{x-2} = \frac{2}{(1-q)^3}$$

$$\sum_{x=1}^{\infty} x^2 q^{x-1} - \sum_{x=1}^{\infty} x q^{x-1} = \frac{2q}{(1-q)^3} \qquad \sum_{x=1}^{\infty} x^2 q^{x-1} = \frac{2q}{(1-q)^3} + \frac{1}{(1-q)^2} = \frac{1+q}{(1-q)^3}$$

$E(X^2) = 0.1\sum_{x=1}^{\infty} x^2 (0.9)^{x-1} = 0.1\dfrac{1+0.9}{(1-0.9)^2} = 190$

$V(X) = E(X^2) – [E(X)]^2 = 190 – 100 = 90$

3-75. Let X denote the number of failed devices. Here X ϵ {0,1,2}

P(X = 0) = 0.8(0.9) = 0.72
P(X = 1) = 0.8(0.1) + 0.2(0.9) = 0.26
P(X = 2) = 0.2(0.1) = 0.02
E(X) = 0(0.72) + 1(0.26) + 2(0.02) = 0.30

Section 3-5

3-76. E(X) = (0 + 99)/2 = 49.5, V(X) = $[(99 – 0 + 1)^2 – 1]/12 = 833.25$

3-77. E(X) = (3 + 1)/2 = 2, V(X) = $[(3 – 1 + 1)^2 – 1]/12 = 0.667$

3-78. X=(1/100)Y, Y = 15, 16, 17, 18, 19.

E(X) = (1/100) E(Y) = $\dfrac{1}{100}\left(\dfrac{15+19}{2}\right) = 0.17$ mm

$$V(X) = \left(\frac{1}{100}\right)^2\left[\frac{(19-15+1)^2-1}{12}\right] = 0.0002 \text{ mm}^2$$

3-79. $E(X) = 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{4}\right) + 4\left(\frac{1}{4}\right) + 5\left(\frac{1}{4}\right) = 3.5$

$$V(X) = (2)^2\left(\frac{1}{4}\right) + (3)^2\left(\frac{1}{4}\right) + (4)^2\left(\frac{1}{4}\right) + (5)^2\left(\frac{1}{4}\right) - (3.5)^2 = \frac{5}{4} = 1.25$$

3-80. X = 590 + 0.1Y, Y = 0, 1, 2, ..., 9

E(X) = $590 + 0.1\left(\dfrac{0+9}{2}\right) = 590.45$ mm

$$V(X) = (0.1)^2\left[\frac{(9-0+1)^2-1}{12}\right] = 0.0825 \text{ mm}^2$$

3-81. a = 675, b = 700
a) μ = E(X) = (a + b)/2 = 687.5
V(X) = $[(b – a + 1)^2 – 1]/12 = 56.25$

b) a = 75, b = 100

$\mu = E(X) = (a + b)/2 = 87.5$
$V(X) = [(b - a + 1)^2 - 1]/12 = 56.25$
The range of values is the same, so the mean shifts by the difference in the two minimums (or maximums) whereas the variance does not change.

3-82. X is a discrete random variable because it denotes the number of fields out of 28 that are in error.
However, X is not uniform because $P(X = 0) \neq P(X = 1)$.

3-83. The range of Y is 0, 5, 10, ..., 45, $E(X) = (0 + 9)/2 = 4.5$
$E(Y) = 0(1/10)+5(1/10)+...+45(1/10)$
$= 5[0(0.1) +1(0.1)+ ... +9(0.1)]$
$= 5E(X)$
$= 5(4.5)$
$= 22.5$
$V(X) = 8.25$, $V(Y) = 5^2(8.25) = 206.25$, $\sigma_Y = 14.36$

3-84.

$$E(cX) = \sum_x cxf(x) = c\sum_x xf(x) = cE(X),$$

$$V(cX) = \sum_x (cx - c\mu)^2 f(x) = c^2 \sum_x (x-\mu)^2 f(x) = cV(X)$$

3-85. $E(X) = (9+5)/2 = 7$, $V(X) = [(9-5+1)^2-1]/12 = 2$, $\sigma = 1.414$

3-86. $f(x_i) = \frac{3\times 10^8}{10^9} = 0.3$

3-87. A = the event that your number is called
$P(A) = 1000/(10^7) = 0.0001$

3-88.
No. The range of X is {0, 1, 2,…, 28} so X is a discrete random variable, but not uniform. For example, $P(X=0) = 0.995^{28}$ is not equal to $P(X=1) = 28(0.995^{27})(0.005)$.

3-89.
No. The range of X is {0, 1, 2,…, 10} so X is a discrete random variable, but not uniform. For example, $P(X=0)$ is not equal to $P(X=1)$.

3-90.
No. X is a discrete random variable but not uniform. For example, the probability that a patient is selected from hospital 1 (= 3820/16814) is different than the probability a patient is selected from hospital 2 (= 5163/16814).

	Hospital				
	1	2	3	4	Total
Total	5292	6991	5640	4329	22,252
LWBS	195	270	246	242	953
Admitted	1277	1558	666	984	4485
Not admitted	3820	5163	4728	3103	16,814

<u>Section 3-6</u>

3-91. A binomial distribution is based on independent trials with two outcomes and a constant probability of success on each trial.

a) reasonable
b) independence assumption not reasonable
c) The probability that the second component fails depends on the failure time of the first component. The binomial distribution is not reasonable.
d) not independent trials with constant probability
e) probability of a correct answer not constant
f) reasonable
g) probability of finding a defect not constant
h) if the fills are independent with a constant probability of an underfill, then the binomial distribution for the number packages underfilled is reasonable
i) because of the bursts, each trial (that consists of sending a bit) is not independent
j) not independent trials with constant probability

3-92. (a) $P(X \le 3) = 0.411$
(b) $P(X>10) = 1 - 0.9994 = 0.0006$
(c) $P(X=6) = 0.1091$
(d) $P(6 \le X \le 11) = 0.9999 - 0.8042 = 0.1957$

3-93. (a) $P(X \le 2) = 0.9298$
(b) $P(X>8) = 0$
(c) $P(X=4) = 0.0112$
(d) $P(5 \le X \le 7) = 1 - 0.9984 = 0.0016$

3-94. a) $P(X = 5) = \binom{10}{5} 0.5^5 (0.5)^5 = 0.2461$

b) $P(X \le 2) = \binom{10}{0} 0.5^0 0.5^{10} + \binom{10}{1} 0.5^1 0.5^9 + \binom{10}{2} 0.5^2 0.5^8$

$= 0.5^{10} + 10(0.5)^{10} + 45(0.5)^{10} = 0.0547$

c) $P(X \ge 9) = \binom{10}{9} 0.5^9 (0.5)^1 + \binom{10}{10} 0.5^{10} (0.5)^0 = 0.0107$

d) $P(3 \le X < 5) = \binom{10}{3} 0.5^3 0.5^7 + \binom{10}{4} 0.5^4 0.5^6$

$= 120(0.5)^{10} + 210(0.5)^{10} = 0.3223$

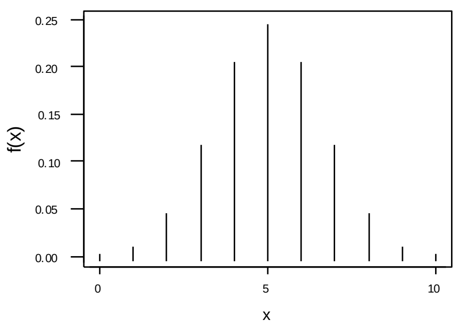
3-95. a) $P(X = 5) = \binom{10}{5} 0.01^5 (0.99)^5 = 2.40 \times 10^{-8}$

$$b) P(X \le 2) = \binom{10}{0} 0.01^0 (0.99)^{10} + \binom{10}{1} 0.01^1 (0.99)^9 + \binom{10}{2} 0.01^2 (0.99)^8$$
$$= 0.9999$$

$$c) P(X \ge 9) = \binom{10}{9} 0.01^9 (0.99)^1 + \binom{10}{10} 0.01^{10} (0.99)^0 = 9.91 \times 10^{-18}$$

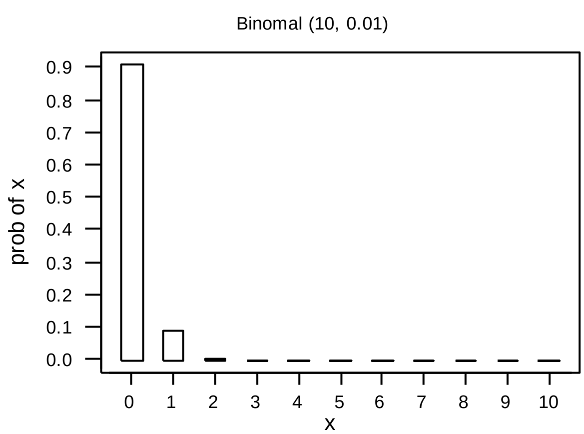
$$d) P(3 \le X < 5) = \binom{10}{3} 0.01^3 (0.99)^7 + \binom{10}{4} 0.01^4 (0.99)^6 = 1.138 \times 10^{-4}$$

3-96.



a) $P(X = 5) = 0.9999$, x= 5 is most likely, also $E(X) = np = 10(0.5) = 5$
b) Values x= 0 and x=10 are the least likely, the extreme values

3-97.



P(X = 0) = 0.904, P(X = 1) = 0.091, P(X = 2) = 0.004, P(X = 3) = 0. P(X = 4) = 0 and so forth.
Distribution is skewed with $E(X) = np = 10(0.01) = 0.1$

a) The most-likely value of X is 0.
b) The least-likely value of X is 10.

3-98. n=3 and p=0.5

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.125 & 0 \le x < 1 \\ 0.5 & 1 \le x < 2 \\ 0.875 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases} \quad \text{where} \quad \begin{aligned} f(0) &= \left(\frac{1}{2}\right)^3 = \frac{1}{8} \\ f(1) &= 3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 = \frac{3}{8} \\ f(2) &= 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) = \frac{3}{8} \\ f(3) &= \left(\frac{1}{4}\right)^3 = \frac{1}{8} \end{aligned}$$

3-99. The binomial distribution has n = 3 and p = 0.25

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.4219 & 0 \le x < 1 \\ 0.8438 & 1 \le x < 2 \\ 0.9844 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases} \quad \text{where} \quad \begin{aligned} f(0) &= \left(\frac{3}{4}\right)^3 = \frac{27}{64} \\ f(1) &= 3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2 = \frac{27}{64} \\ f(2) &= 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) = \frac{9}{64} \\ f(3) &= \left(\frac{1}{4}\right)^3 = \frac{1}{64} \end{aligned}$$

3-100. Let X denote the number of defective circuits.
Then, X has a binomial distribution with n = 40 and p = 0.01
P(X = 0) = $\binom{40}{0}0.01^0 0.99^{40} = 0.6690$.

3-101. Let X denote the number of times the line is occupied.
Then, X has a binomial distribution with n = 10 and p = 0.4

a) $P(X = 3) = \binom{10}{3}0.4^3(0.6)^7 = 0.215$

b) Let Z denote the number of time the line is NOT occupied.
Then Z has a binomial distribution with n =10 and p = 0.6.

$$P(Z \ge 1) = 1 - P(Z = 0) = 1 - \binom{10}{0}0.6^0 0.4^{10} = 0.9999$$

c) $E(X) = 10(0.4) = 4$

3-102. Let X denote the number of questions answered correctly.
Then, X is binomial with n = 25 and p = 0.25.

$$a) P(X > 20) = \binom{25}{21}0.25^{21}(0.75)^4 + \binom{25}{22}0.25^{22}(0.75)^3$$
$$+ \binom{25}{23}0.25^{23}(0.75)^2 + \binom{25}{24}0.25^{24}(0.75)^1 + \binom{25}{25}0.25^{25}(0.75)^0 = 9.677 \times 10^{-10}$$

$$b) P(X < 5) = \binom{25}{0}0.25^0(0.75)^{25} + \binom{25}{1}0.25^1(0.75)^{24} + \binom{25}{2}0.25^2(0.75)^{23}$$
$$+ \binom{25}{3}0.25^3(0.75)^{22} + \binom{25}{4}0.25^4(0.75)^{21} = 0.2137$$

3-103. Let X denote the number of mornings the light is green.

$$a)\quad P(X=1)=\binom{5}{1}0.2^1 0.8^4 = 0.410$$

$$b)\quad P(X=4)=\binom{20}{4}0.2^4 0.8^{16} = 0.218$$

$$c)\quad P(X>4)=1-P(X\le 4)=1-0.630=0.370$$

3-104. X = number of samples mutated
X has a binomial distribution with p=0.01, n=15

(a) P(X=0) = $\binom{15}{0}p^0(1-p)^{15}$ = 0.86

(b) P(X≤1)=P(X=0)+P(X=1)= 0.99

(c) P(X>7)=P(X=8)+P(X=9)+…+P(X=15)= 0

3-105. (a) n = 20, p = 0.6122,
P(X≥1) = 1 - P(X=0) = 1

(b)P(X≥3) = 1 - P(X<3)= 0.999997

(c) μ = E(X) = np = 20(0.6122) = 12.244

V(X) = np(1 - p) = 4.748

σ = $\sqrt{V(X)}$ =2.179

3-106. The binomial distribution has n = 20 and p = 0.13

(a) P(X = 3) = $\binom{20}{3}p^3(1-p)^{17}$ =0.235

(b) P(X ≥ 3) = 1 - P(X<3)=0.492

(c) μ = E(X) = np = 20(0.13) = 2.6

V(X) = np(1 - p) = 2.262

σ = $\sqrt{V(X)}$ = 1.504

3-107. (a) Binomial distribution, p =$10^4/36^9$ = 4.59394E-06, n = 1E09

(b) P(X=0) = $\binom{1E09}{0}p^0(1-p)^{1E09}$ = 0

(c) μ = E(X) = np =1E09(0.45939E-06) = 4593.9

V(X) = np(1 - p) = 4593.9

3-108. E(X) = 20(0.01) = 0.2
V(X) = 20(0.01) (0.99) = 0.198

$$\mu_X + 3\sigma_X = 0.2 + 3\sqrt{0.198} = 1.53$$

a) X is binomial with n = 20 and p = 0.01

$$P(X > 1.53) = P(X \geq 2) = 1 - P(X \leq 1)$$
$$= 1 - \left[\binom{20}{0} 0.01^0 0.99^{20} + \binom{20}{1} 0.01^1 0.99^{19}\right] = 0.0169$$

b) X is binomial with n = 20 and p = 0.04

$$P(X > 1) = 1 - P(X \leq 1)$$
$$= 1 - \left[\binom{20}{0} 0.04^0 0.96^{20} + \binom{20}{1} 0.04^1 0.96^{19}\right] = 0.1897$$

c) Let Y denote the number of times X exceeds 1 in the next five samples.
Then, Y is binomial with n = 5 and p = 0.190 from part b.

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \left[\binom{5}{0} 0.190^0 0.810^5\right] = 0.651$$

The probability is 0.651 that at least one sample from the next five will contain more than one defective

3-109. Let X denote the passengers with tickets that do not show up for the flight.
Then, X has a binomial distribution with n = 125 and p = 0.1

$$a) P(X \geq 5) = 1 - P(X \leq 4)$$
$$= 1 - \left[\begin{array}{l} \binom{125}{0} 0.1^0 (0.9)^{125} + \binom{125}{1} 0.1^1 (0.9)^{124} + \binom{125}{2} 0.1^2 (0.9)^{123} \\ + \binom{125}{3} 0.1^3 (0.9)^{122} + \binom{125}{4} 0.1^4 (0.9)^{121} \end{array}\right]$$
$$= 0.9961$$

$$b) P(X > 5) = 1 - P(X \leq 5) = 0.9886$$

3-110. Let X denote the number of defective components among those stocked.

$$a) \quad P(X = 0) = \binom{100}{0} 0.02^0 0.98^{100} = 0.133$$

$$b) \quad P(X \leq 2) = \binom{102}{0} 0.02^0 0.98^{102} + \binom{102}{1} 0.02^1 0.98^{101} + \binom{102}{2} 0.02^2 0.98^{100} = 0.666$$

$$c) \quad P(X \leq 5) = 0.981$$

3-111. P(length of stay ≤ 4) = 0.516

a) Let *N* denote the number of people (out of five) that wait less than or equal to 4 hours.

$$P(N = 1) = \binom{5}{1}(0.516)^1(0.484)^4 = 0.142$$

b) Let *N* denote the number of people (out of five) that wait more than 4 hours.

$$P(N = 2) = \binom{5}{2}(0.484)^2(0.516)^3 = 0.322$$

c) Let *N* denote the number of people (out of five) that wait more than 4 hours.

$$P(N \geq 1) = 1 - P(N = 0) = 1 - \binom{5}{0}(0.516)^5(0.484)^0 = 0.963$$

3-112. Probability a person leaves without being seen (LWBS) = 195/5292 = 0.037

a) $P(X = 1) = \binom{4}{1}(0.037)^1(0.963)^3 = 0.132$

b) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)]$

$$= 1 - \binom{4}{0}(0.037)^0(0.963)^4 - \binom{4}{1}(0.037)^1(0.963)^3 = 0.008$$

c) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.86 = 0.14$

3-113. $P(change < 4\ days) =$ 0.3. Let *X* = number of the 10 changes made in less than 4 days.

a) $P(X=7)=\binom{10}{7}(0.3)^7(0.7)^3=0.009$

b) $P(X\le 2)=P(X=0)+P(X=1)+P(X=2)$

$=\binom{10}{0}(0.3)^0(0.7)^{10}+\binom{10}{1}(0.3)^1(0.7)^9+\binom{10}{2}(0.3)^2(0.7)^8$

$=0.028+0.121+0.233=0.382$

c) $P(X\ge 1)=1-P(X=0)=1-\binom{10}{0}(0.3)^0(0.7)^{10}=1-0.028=0.972$

d) $E(X)=np=10(0.3)=3$

3-114. $P(reaction<272K)=\ 0.54$

a) $P(X=12)=\binom{20}{12}(0.54)^{12}(0.46)^8=0.155$

b) $P(X\ge 19)=P(X=19)+P(X=20)$

$$=\binom{20}{19}(0.54)^{19}(0.46)^1+\binom{20}{20}(0.54)^{20}(0.46)^0=0.00008$$

c) $P(X\ge 18)=P(X=18)+P(X=19)+P(X=20)$

$$=\binom{20}{18}(0.54)^{18}(0.46)^2+0.00008=0.00069$$

d) $E(X)=np=20(0.54)=10.8$

3-115.

Let X = the number of visitors that provide contact data. Then X is a binomial random variable with p = 0.01 and n = 1000.

a) $\mathrm{P(X=0)}=\binom{1000}{0}0.01^0(1-0.01)^{1000}\cong 0$

b) $\mathrm{P(X=10)}=\binom{1000}{10}0.01^{10}(1-0.01)^{1000-10}\cong 0.126$

c) $\mathrm{P(X>3)}=1-[\mathrm{P(X=0)+P(X=1)+P(X=2)+P(X=3)}]$

$\mathrm{P(X>3)}=1-[0+0.0004+0.0022+0.0074]\cong 0.99$

3-116.

Let X = the number of device failures. Then X is a binomial random variable with p = 0.05 and n = 2. Therefore the probability mass function is

$\mathrm{P(X=x)}=\binom{2}{x}0.05^x(0.95)^{2-x}$

$\mathrm{P(X=0)}=0.95^2=0.9025$

$\mathrm{P(X=1)}=2(0.05)0.95=0.095$

$\mathrm{P(X=2)}=0.05^2=0.0025$

The binomial distribution does not apply to the number of failures in the example because the probability of failure is not the same for both devices.

3-117.

Let X = the number of cameras failing. Then X is a binomial random variable with probability of failing p = 0.2 and n is to be determined.

We need to find the smallest n such that $\mathrm{P(X\ge 1)}\ge 0.95$

Equivalently, $1-\mathrm{P(X=0)}\ge 0.95$ or $\mathrm{P(X=0)}\le 0.05$

$0.8^n\le 0.05$, n ln(0.8) $\le$ ln(0.05), n=ln(0.05)/ln(0.8) = 13.4

Therefore, the smallest sample size n that satisfies the condition is 14.

3-118.

Let X = the number of patients who are LWBS from hospital 4. Then X is a binomial random variable with *p* = 242/4329 and *n* is to be calculated.

We need to find the smallest n such that $P(X \geq 1) \geq 0.90$

Equivalently, $1 - P(X = 0) \geq 0.90$ or $P(X = 0) \leq 0.1$

$\left(1 - \frac{242}{4329}\right)^n \leq 0.1$, $n\ln(0.944) \leq \ln(0.1)$, $n = \ln(0.1)/\ln(0.944) = 39.95$

Therefore, the smallest sample size that satisfies the condition is *n* = 40.

Section 3-7

3-119.

a) $P(X = 1) = (1-0.5)^0 0.5 = 0.5$

b) $P(X = 4) = (1-0.5)^3 0.5 = 0.5^4 = 0.0625$

c) $P(X = 8) = (1-0.5)^7 0.5 = 0.5^8 = 0.0039$

d) $P(X \leq 2) = P(X = 1) + P(X = 2) = (1-0.5)^0 0.5 + (1-0.5)^1 0.5$

$= 0.5 + 0.5^2 = 0.75$

e) $P(X > 2) = 1 - P(X \leq 2) = 1 - 0.75 = 0.25$

3-120. E(X) = 2.5 = 1/p so that p = 0.4

a) $P(X = 1) = (1-0.4)^0 0.4 = 0.4$

b) $P(X = 4) = (1-0.4)^3 0.4 = 0.0864$

c) $P(X = 5) = (1-0.5)^4 0.5 = 0.05184$

d) $P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3)$

$= (1-0.4)^0 0.4 + (1-0.4)^1 0.4 + (1-0.4)^2 0.4 = 0.7840$

e) $P(X > 3) = 1 - P(X \leq 3) = 1 - 0.7840 = 0.2160$

3-121. Let X denote the number of trials to obtain the first success.

a) E(X) = 1/0.2 = 5

b) Because of the lack of memory property, the expected value is still 5.

3-122. a) E(X) = 4/0.2 = 20

b) P(X = 20) = $\binom{19}{3}(0.80)^{16}0.2^4 = 0.0436$

c) P(X = 19) = $\binom{18}{3}(0.80)^{15}0.2^4 = 0.0459$

d) P(X = 21) = $\binom{20}{3}(0.80)^{17}0.2^4 = 0.0411$

e) The most likely value for X should be near μ = 20. By trying several cases, the most likely value is x = 19.

3-123. Let X denote the number of trials to obtain the first successful alignment.

Then X is a geometric random variable with p = 0.8

a) $P(X = 4) = (1-0.8)^3 0.8 = 0.2^3 0.8 = 0.0064$

b) $P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$$= (1-0.8)^0 0.8 + (1-0.8)^1 0.8 + (1-0.8)^2 0.8 + (1-0.8)^3 0.8$$
$$= 0.8 + 0.2(0.8) + 0.2^2(0.8) + 0.2^3 0.8 = 0.9984$$

c) $P(X \geq 4) = 1 - P(X \leq 3) = 1 - [P(X = 1) + P(X = 2) + P(X = 3)]$
$$= 1 - [(1-0.8)^0 0.8 + (1-0.8)^1 0.8 + (1-0.8)^2 0.8]$$
$$= 1 - [0.8 + 0.2(0.8) + 0.2^2(0.8)] = 1 - 0.992 = 0.008$$

3-124.

X = the number of people tested to detect two with the gene, X ϵ {2, 3, 4, ...}. Then X has a negative binomial distribution with p = 0.1 and r = 2. We have to find P(X >= 4).

a) $P(X \geq 4) = 1 - [P(X = 2) + P(X = 3)]$ where $P(X = x) = \binom{x-1}{2-1}(1-0.1)^{x-2} 0.1^2$

$= 1 - [0.01 + 0.018] = 0.972$

b) $E[X] = \dfrac{r}{p} = \dfrac{2}{0.1} = 20$

3-125. Let X denote the number of calls needed to obtain a connection.
Then, X is a geometric random variable with p = 0.02.

a) $P(X = 10) = (1-0.02)^9 0.02 = 0.98^9 0.02 = 0.0167$

b) $P(X > 5) = 1 - P(X \leq 4) = 1 - [P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)]$
$$= 1 - [0.02 + 0.98(0.02) + 0.98^2(0.02) + 0.98^3(0.02) + 0.98^3(0.02) + 0.98^4(0.02)]$$
$= 1 - 0.0961 = 0.9039$

May also use the fact that P(X > 5) is the probability of no connections in 5 trials. That is,

$P(X > 5) = \binom{5}{0} 0.02^0 0.98^5 = 0.9039$

c) E(X) = 1/0.02 = 50

3-126. X = number of opponents until the player is defeated.
p=0.8, the probability of the opponent defeating the player.
(a) $f(x) = (1 - p)^{x-1}p = 0.8^{(x-1)}(0.2)$
(b) P(X>2) = 1 – P(X = 1) – P(X = 2) = 0.64
(c) μ = E(X) = 1/p = 5
(d) P(X ≥ 4) = 1 - P(X = 1) - P(X = 2) - P(X = 3) = 0.512
(e) The probability that a player contests four or more opponents is obtained in part (d), which is p_o = 0.512.
Let Y represent the number of game plays until a player contests four or more opponents.
Then, $f(y) = (1 - p_o)^{y-1}p_o$.
$\mu_Y = E(Y) = 1/p_o = 1.95$

3-127. p = 0.13
(a) $P(X = 1) = (1 - 0.13)^{1-1}(0.13) = 0.13$
(b) $P(X = 3) = (1 - 0.13)^{3-1}(0.13) = 0.098$
(c) μ = E(X) = 1/p = 7.69 ≈ 8

3-128. X = number of attempts before the hacker selects a user password.
(a) $p = 9900/36^6 = 0.0000045$
μ = E(X) = 1/p = 219877
$V(X) = (1 - p)/p^2 = 4.938E10^{10}$
$\sigma = \sqrt{V(X)}$ = 222,222
(b) $p = 100/36^3 = 0.00214$
μ = E(X) = 1/p = 467

V(X)= $(1 - p)/p^2$ = 217892.39

$\sigma = \sqrt{V(X)}$ = 466.78

Based on the answers to (a) and (b) above, it is clearly more secure to use a 6 character password.

3-129. p = 0.005 and r = 8

a) $P(X = 8) = 0.005^8 = 3.91E10^{-19}$

b) $\mu = E(X) = \dfrac{1}{0.005} = 200$ days

c) Mean number of days until all 8 computers fail. Now we use $p=3.91\text{x}10^{-19}$

$\mu = E(Y) = \dfrac{1}{3.91x10^{-91}} = 2.56x10^{18}$ days or 7.01 $\text{x}10^{15}$ years

3-130. Let Y denote the number of samples needed to exceed 1 in Exercise 3-66.
Then Y has a geometric distribution with p = 0.0169.

a) $P(Y = 10) = (1 - 0.0169)^9(0.0169) = 0.0145$

b) Y is a geometric random variable with p = 0.1897 from Exercise 3-66.

$P(Y = 10) = (1 - 0.1897)^9(0.1897) = 0.0286$

c) E(Y) = 1/0.1897 = 5.27

3-131. Let X denote the number of transactions until all computers have failed.
Then, X is negative binomial random variable with $p = 10^{-8}$ and r = 3.

a) $E(X) = 3 \text{ x } 10^8$

b) $V(X) = [3(1-10^{-8}0]/(10^{-16}) = 3.0 \text{ x } 10^{16}$

3-132. (a) $p^6 = 0.6$, p = 0.918

(b) $0.6p^2 = 0.4$, p = 0.816

3-133. Negative binomial random variable f(x; p, r) = $\binom{x-1}{r-1}(1-p)^{x-r}p^r$

When r = 1, this reduces to $f(x) = (1-p)^{x-1}p$, which is the pdf of a geometric random variable.
Also, E(X) = r/p and $V(X) = [r(1-p)]/p^2$ reduce to E(X) = 1/p and $V(X) = (1-p)/p^2$, respectively.

3-134. $P(reaction < 272K) = 0.54$

a) $P(X = 10) = 0.46^9 0.54^1 = 0.0005$

b) $\mu = E(X) = \frac{1}{p} = \frac{1}{0.54} = 1.85$

c) $P(X \leq 3) = P(X = 3) + P(X = 2) + P(X = 1)$

$= 0.46^2 0.54^1 + 0.46^1 0.54^1 + 0.46^0 0.54^1 = 0.903$

d) $\mu = E(X) = \frac{r}{p} = \frac{2}{0.54} = 3.70$

3-135. a) Probability that color printer will be discounted = 1/10 = 0.01

$\mu = E(X) = \frac{1}{p} = \frac{1}{0.10} = 10$ days

b) $P(X = 10) = 0.9^9 0.1 = 0.039$

c) Lack of memory property implies the answer equals $P(X = 10) = 0.9^9 0.1 = 0.039$

d) $P(X \leq 3) = P(X = 3) + P(X = 2) + P(X = 1) = 0.9^2 0.1 + 0.9^1 0.1 + 0.1 = 0.271$

3-136. $P(LWBS) = 0.037$

a) $P(X = 5) = 0.963^4 0.037^1 = 0.032$

b) $P(X = 5) + P(X = 6) = 0.963^4 0.037^1 + 0.963^5 0.037^1 = 0.062$

c) $P(X \leq 4) = P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1)$
$= 0.963^3 0.037^1 + 0.963^2 0.037^1 + 0.963^1 0.037^1 + 0.037 = 0.140$

d) $\mu = E(X) = \frac{r}{p} = \frac{3}{0.037} = 81.08$

3-137.

X = the number of cameras tested to detect two failures, X ϵ {2, 3, 4, ...}. Then X has a negative binomial distribution with p = 0.2 and r = 2.
Y = the number of cameras tested to detect three failures, Y ϵ {3, 4, 5, ...}. Then Y has a negative binomial distribution with p = 0.2 and r = 3.
Note that the events are described in terms of the number of failures, so *p* = 1-0.8 =0.2.

a) $P(X = 10) = \binom{10-1}{2-1}(1-0.2)^{10-2}0.2^2 = 0.0604$

b) $P(X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = 0.04 + 0.064 + 0.0768 = 0.1808$ where

$$P(X = 2) = \binom{2-1}{2-1}(1-0.2)^{2-2}0.2^2 = 0.04$$

$$P(X = 3) = \binom{3-1}{2-1}(1-0.2)^{3-2}0.2^2 = 0.064$$

$$P(X = 4) = \binom{4-1}{2-1}(1-0.2)^{4-2}0.2^2 = 0.0768$$

c) $E[Y] = \frac{r}{p} = \frac{3}{0.2} = 15$

3-138.

X = the number of defective bulbs in an array of 30 LED bulbs. Here X is a binomial random variable with *p* = 0.001 and *n* = 30

a) $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$

$$P(X \geq 2) = 1 - \left[\binom{30}{0}0.001^0(1-0.001)^{30} + \binom{30}{1}0.001^1(1-0.001)^{29}\right] = 0.0004$$

b) Let Y = number of automotive lights tested to obtain one light with two or more defective bulbs among thirty LED bulbs. Here Y is distributed with negative binomial with success probability *p* = 0.0004 and *r* = 1.

$$E[Y] = \frac{r}{p} = \frac{1}{0.0004} = 2500$$

3-139.

X = the number of patients selected from hospital 4 in order to admit 2. Then X has a negative binomial distribution with p = 0.23 and r =2.
Y = the number of patients selected from hospital 4 in order to admit 10. Then Y has a negative binomial distribution with p = 0.23 and r =10.

a) $= \frac{984}{4329} = 0.2273$

b) $P(X \leq 4) = P(X = 4) + P(X = 3) + P(X = 2)$

$$P(X = 4) = \binom{4-1}{2-1}(1-0.23)^{4-2}0.23^2 = 0.094$$

$$\mathrm{P(X=3)}=\binom{3-1}{2-1}(1-0.23)^{3-2}0.23^2=0.062$$

$$\mathrm{P(X=2)}=\binom{2-1}{2-1}(1-0.23)^{2-2}0.23^2=0.031$$

$$\mathrm{P(X\le 4)}=0.094+0.062+0.031=0.185$$

c) $E[Y]=\dfrac{r}{p}=\dfrac{10}{0.227}\cong 43.99$

3-140. Let X denote the number of customers who visit the website to obtain the first order.

a) Yes, since the customers behave independently and the probability of a success (i.e., obtaining an order) is the same for all customers.
b) A = the event that a customer views five or fewer pages
B = the event that the customer orders

$$P(B)=P(B\cap A)+P(B\cap A')=P(B\,|\,A)P(A)+P(B\,|\,A')P(A')$$

$$P(B)=0.01(1-0.25)+0.1(0.25)=0.0325$$

Then X has a geometric distribution with p = 0.0325
P(X = 10) = 0.0325 (1 - 0.0325)^9 = 0.024

Section 3-8

3-141. X has a hypergeometric distribution with N = 100, n = 4, K = 20

a) $P(X=1)=\dfrac{\binom{20}{1}\binom{80}{3}}{\binom{100}{4}}=\dfrac{20(82160)}{3921225}=0.4191$

b) $P(X=6)=0$, the sample size is only 4

c) $P(X=4)=\dfrac{\binom{20}{4}\binom{80}{0}}{\binom{100}{4}}=\dfrac{4845(1)}{3921225}=0.001236$

d) $E(X)=np=n\dfrac{K}{N}=4\left(\dfrac{20}{100}\right)=0.8$

$$V(X)=np(1-p)\left(\frac{N-n}{N-1}\right)=4(0.2)(0.8)\left(\frac{96}{99}\right)=0.6206$$

3-142. a) $P(X=1)=\dfrac{\binom{4}{1}\binom{16}{3}}{\binom{20}{4}}=\dfrac{(4\times16\times15\times14)/6}{(20\times19\times18\times17)/24}=0.4623$

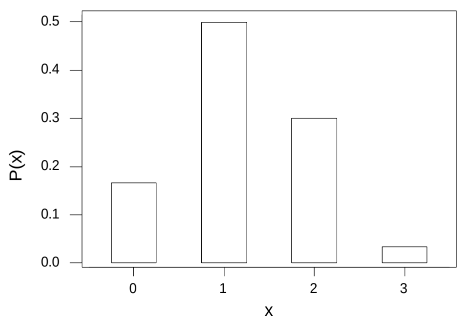
b) $P(X=4)=\dfrac{\binom{4}{4}\binom{16}{0}}{\binom{20}{4}}=\dfrac{1}{(20\times19\times18\times17)/24}=0.00021$

c)

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= \frac{\binom{4}{0}\binom{16}{4}}{\binom{20}{4}} + \frac{\binom{4}{1}\binom{16}{3}}{\binom{20}{4}} + \frac{\binom{4}{2}\binom{16}{2}}{\binom{20}{4}}$$
$$= \frac{\left(\frac{16\times15\times14\times13}{24} + \frac{4\times16\times15\times14}{6} + \frac{6\times16\times15}{2}\right)}{\left(\frac{20\times19\times18\times17}{24}\right)} = 0.9866$$

d) E(X) = 4(4/20) = 0.8
V(X) = 4(0.2)(0.8)(16/19) = 0.539

3-143. Here N = 10, n = 3, K= 4



3-144. (a) f(x) = $\binom{24}{x}\binom{12}{3-x} / \binom{36}{3}$

(b) μ = E(X) = np = 3*24/36=2
V(X) = np(1-p)(N-n)/(N-1) = 2(1 - 24/36)(36 - 3)/(36 - 1) = 0.629

(c) P(X≤2) =1 - P(X=3) = 0.717

3-145. Let X denote the number of men who carry the marker on the male chromosome for an increased risk for high blood pressure. Here N = 800, K = 240
a) n = 10
$$P(X = 1) = \frac{\binom{240}{1}\binom{560}{9}}{\binom{800}{10}} = \frac{\left(\frac{240!}{1!239!}\right)\left(\frac{560!}{9!551!}\right)}{\frac{800!}{10!790!}} = 0.1201$$
b) n = 10
$$P(X > 1) = 1 - P(X \le 1) = 1 - [P(X = 0) + P(X = 1)]$$
$$P(X = 0) = \frac{\binom{240}{0}\binom{560}{10}}{\binom{800}{10}} = \frac{\left(\frac{240!}{0!240!}\right)\left(\frac{560!}{10!550!}\right)}{\frac{800!}{10!790!}} = 0.0276$$
$$P(X > 1) = 1 - P(X \le 1) = 1 - [0.0276 + 0.1201] = 0.8523$$

3-146. Let X denote the number of cards in the sample that are defective.
a)

$$P(X \geq 1) = 1 - P(X = 0)$$

$$P(X = 0) = \frac{\binom{20}{0}\binom{120}{20}}{\binom{140}{20}} = \frac{\frac{120!}{20!100!}}{\frac{140!}{20!120!}} = 0.0356$$

$$P(X \geq 1) = 1 - 0.0356 = 0.9644$$

b)

$$P(X \geq 1) = 1 - P(X = 0)$$

$$P(X = 0) = \frac{\binom{5}{0}\binom{135}{20}}{\binom{140}{20}} = \frac{\frac{135!}{20!115!}}{\frac{140!}{20!120!}} = \frac{135!120!}{115!140!} = 0.4571$$

$$P(X \geq 1) = 1 - 0.4571 = 0.5429$$

3-147. N = 300

(a) K = 243, n = 3, P(X = 1)=0.087

(b) P(X ≥ 1) = 0.9934

(c) K = 26 + 13 = 39, P(X = 1)=0.297

(d) K = 300 - 18 = 282

P(X ≥ 1) = 0.9998

3-148. Let X denote the count of the numbers in the state's sample that match those in the player's sample. Then, X has a hypergeometric distribution with N = 40, n = 6, and K = 6.

a) $P(X = 6) = \frac{\binom{6}{6}\binom{34}{0}}{\binom{40}{6}} = \left(\frac{40!}{6!34!}\right)^{-1} = 2.61 \times 10^{-7}$

b) $P(X = 5) = \frac{\binom{6}{5}\binom{34}{1}}{\binom{40}{6}} = \frac{6 \times 34}{\binom{40}{6}} = 5.31 \times 10^{-5}$

c) $P(X = 4) = \frac{\binom{6}{4}\binom{34}{2}}{\binom{40}{6}} = 0.00219$

d) Let Y denote the number of weeks needed to match all six numbers.

Then, Y has a geometric distribution with p = $\frac{1}{3{,}838{,}380}$ and

E(Y) = 1/p = 3,838,380 weeks. This is more than 738 centuries!

3-149. Let X denote the number of blades in the sample that are dull.

a)

$$P(X \geq 1) = 1 - P(X = 0)$$

$$P(X = 0) = \frac{\binom{10}{0}\binom{38}{5}}{\binom{48}{5}} = \frac{\frac{38!}{5!33!}}{\frac{48!}{5!43!}} = \frac{38!43!}{48!33!} = 0.2931$$

$$P(X \geq 1) = 1 - P(X = 0) = 0.7069$$

b) Let Y denote the number of days needed to replace the assembly.

P(Y = 3) = $0.2931^2(0.7069) = 0.0607$

c) On the first day, $P(X = 0) = \frac{\binom{2}{0}\binom{46}{5}}{\binom{48}{5}} = \frac{\frac{46!}{5!41!}}{\frac{48!}{5!43!}} = \frac{46!43!}{48!41!} = 0.8005$

On the second day, $P(X=0)=\frac{\binom{6}{0}\binom{42}{5}}{\binom{48}{5}}=\frac{\frac{42!}{5!37!}}{\frac{48!}{5!43!}}=\frac{42!43!}{48!37!}=0.4968$

On the third day, P(X = 0) = 0.2931 from part a). Therefore,
P(Y = 3) = 0.8005(0.4968)(1-0.2931) = 0.2811.

3-150.

a) For the first exercise, the finite population correction is 96/99.
For the second exercise, the finite population correction is 16/19.
Because the finite population correction for the first exercise is closer to one, the binomial approximation to the distribution of X should be better in that exercise.

b) Assuming X has a binomial distribution with n = 4 and p = 0.2

$$P(X=1)=\binom{4}{1}0.2^1 0.8^3=0.4096$$

$$P(X=4)=\binom{4}{4}0.2^4 0.8^0=0.0016$$

The results from the binomial approximation are close to the probabilities obtained from the hypergeometric distribution.

c) Assume X has a binomial distribution with n = 4 and p = 0.2. Consequently, P(X = 1) and P(X = 4) are the same as computed in part (b) of this exercise. This binomial approximation is not as close to the true answer from the hypergeometric distribution as the results obtained in part (b).

d) X is approximately binomially distributed with n = 20 and p = 20/140 = 1/7.

$$P(X\geq 1)=1-P(X=0)=\binom{20}{0}\left(\tfrac{1}{7}\right)^0\left(\tfrac{6}{7}\right)^{20}=1-0.0458=0.9542$$

The finite population correction is 120/139 = 0.8633

X is approximately binomially distributed with n = 20 and p = 5/140 =1/28

$$P(X\geq 1)=1-P(X=0)=\binom{20}{0}\left(\tfrac{1}{28}\right)^0\left(\tfrac{27}{28}\right)^{20}=1-0.4832=0.5168$$

The finite population correction is 120/139 = 0.8633

3-151. a) $P(X=4)=\frac{\binom{242}{4}\binom{953-242}{0}}{\binom{953}{4}}=0.0041$

b) $P(X=0)=\frac{\binom{242}{0}\binom{953-242}{4}}{\binom{953}{4}}=0.3091$

c) Probability that all visits are from hospital 1

$$P(X=4)=\frac{\binom{195}{4}\binom{953-195}{0}}{\binom{953}{4}}=0.0017$$

Probability that all visits are from hospital 2

$$P(X=4)=\frac{\binom{270}{4}\binom{953-270}{0}}{\binom{953}{4}}=0.0063$$

Probability that all visits are from hospital 3

$$P(X=4)=\frac{\binom{246}{4}\binom{953-246}{0}}{\binom{953}{4}}=0.0044$$

Probability that all visits are from hospital 4

$$P(X=4)=\frac{\binom{242}{4}\binom{953-242}{0}}{\binom{953}{4}}=0.0041$$

Probability that all visits are from the same hospital

$$= .0017 + .0063 + .0044 + .0041 = 0.0165$$

3-152. a) $P(X = 2) = \dfrac{\binom{3290}{2}\binom{7726-3290}{4-2}}{\binom{7726}{4}} = 0.359$

b) $P(X \geq 1) = 1 - P(X = 0) = 1 - \dfrac{\binom{3290}{0}\binom{7726-3290}{4-0}}{\binom{7726}{4}} = 1 - 0.109 = 0.891$

c) $\mu = E(X) = np = 4\left(\frac{3290}{7726}\right) = 1.703$

3-153.

a) Let X_1 denote the number of wafers in the sample with high contamination. Here X_1 has a hypergeometric distribution with N = 940, K = 358, and n = 10 when the sample size is 10.

$$P(X_1 = 4) = \frac{\binom{358}{4}\binom{940-358}{10-4}}{\binom{940}{10}} = 0.25$$

b) Let X_2 denote the number of wafers in the sample with high contamination and from the center of the sputtering tool. Here X_2 has a hypergeometric distribution with N = 940, K = 112, and n = 10 when the sample size is 10.

$$P(X_2 \geq 1) = 1 - P(X_2 = 0) = 1 - \frac{\binom{112}{0}\binom{940-112}{10-0}}{\binom{940}{10}} = 1 - 0.28 = 0.72$$

c) Let X_3 denote the number of wafers in the sample with high contamination or from the edge of the sputtering tool. Here X_3 has a hypergeometric distribution with N = 940, K = 426, and n = 10 when the sample size is 10.

$$P(X_3 = 3) = \frac{\binom{426}{3}\binom{940-426}{10-3}}{\binom{940}{10}} = 0.16$$

where 68 + 112 + 246 = 426

d) Let X_4 denote the number of wafers in the sample with high contamination. Here X_4 has a hypergeometric distribution with N = 940 and K = 358.

Find the minimum n that satisfies the condition $P(X_4 \geq 1) \geq 0.9$

$$P(X_4 \geq 1) = 1 - P(X_4 = 0) = 1 - \frac{\binom{358}{0}\binom{940-358}{n-0}}{\binom{940}{n}} \geq 0.9$$

Through trial of values for n, the minimum n is 5.

3-154.

Let X denote the number of patients in the sample that adhere. Here X has a hypergeometric distribution with N = 500, K = 50 and n = 20 when the sample size is 20.

a) $P(X = 2) = \dfrac{\binom{50}{2}\binom{500-50}{20-2}}{\binom{500}{20}} = 0.291$

b) $P(X < 2) = P(X = 0) + P(X = 1)$

$$= \frac{\binom{50}{0}\binom{500-50}{20-0}}{\binom{500}{20}} + \frac{\binom{50}{1}\binom{500-50}{20-1}}{\binom{500}{20}} = 0.116 + 0.270 = 0.386$$

c) $P(X > 2) = 1 - \left[P(X = 0) + P(X = 1) + P(X = 2)\right] = 1 - \left[0.116 + 0.270 + 0.291\right] = 0.323$

d) $E\left[X\right] = np = n\dfrac{K}{N} = 20\dfrac{50}{500} = 2$

$$Var(X) = np(1-p)\left(\frac{N-n}{N-1}\right) = 20(0.1)(0.9)\left(\frac{480}{499}\right) = 1.73$$

3-155.

Let X = the number of sites with lesions in the sample. Here X has hypergeometric distribution with $N = 50$, $K = 5$ and $n = 8$ when the sample size is 8.

a) $P(X \geq 1) = 1 - P(X = 0) = 1 - \dfrac{\binom{5}{0}\binom{50-5}{8-0}}{\binom{50}{8}} = 0.599$

b) $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - \dfrac{\binom{5}{0}\binom{50-5}{8-0}}{\binom{50}{8}} - \dfrac{\binom{5}{1}\binom{50-5}{8-1}}{\binom{50}{8}} = 0.176$

c) We need to find the minimum n that satisfies the condition $P(X \geq 1) \geq 0.90$

Equivalently, $1 - P(X = 0) \geq 0.90$ or $P(X = 0) \leq 0.10$

$$\frac{\binom{5}{0}\binom{50-5}{n-0}}{\binom{50}{n}} \leq 0.10$$

From trials of n values $\dfrac{\binom{5}{0}\binom{50-5}{18-0}}{\binom{50}{18}} < 0.1 < \dfrac{\binom{5}{0}\binom{50-5}{17-0}}{\binom{50}{17}}$.

The smallest sample size that satisfies the condition is $n = 18$

3-156.

Let *X* = the number of major customers that accept the plan in the sample. Here *X* has hypergeometric distribution with *N* = 50, *K* = 15, and *n* = 10 when the sample size is 10.

a) $P(X = 2) = \dfrac{\binom{15}{2}\binom{50-15}{10-2}}{\binom{50}{10}} = 0.241$

b) $P(X \ge 1) = 1 - P(X = 0) = 1 - \dfrac{\binom{15}{0}\binom{50-15}{10-0}}{\binom{50}{10}} = 0.982$

c) We need to find the minimum *K* that satisfies the condition $P(X \ge 1) \ge 0.95$

Equivalently, $1 - P(X = 0) \ge 0.95$ and $P(X = 0) \le 0.05$. This requires

$$\frac{\binom{K}{0}\binom{50-K}{10-0}}{\binom{50}{10}} \le 0.05$$

We have $\dfrac{\binom{4}{0}\binom{50-4}{10-0}}{\binom{50}{10}} < 0.05 < \dfrac{\binom{3}{0}\binom{50-3}{10-0}}{\binom{50}{10}}$.

The minimum number of major customers that would need to accept the plan to meet the given objective is 4.

Section 3-9

3-157. a) $P(X = 0) = \dfrac{e^{-4}4^0}{0!} = e^{-4} = 0.0183$

b) $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= e^{-4} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!}$$
$$= 0.2381$$

c) $P(X = 4) = \dfrac{e^{-4}4^4}{4!} = 0.1954$

d) $P(X = 8) = \dfrac{e^{-4}4^8}{8!} = 0.0298$

3-158. a) $P(X = 0) = e^{-0.4} = 0.6703$

b) $P(X \le 2) = e^{-0.4} + \dfrac{e^{-0.4}(0.4)}{1!} + \dfrac{e^{-0.4}(0.4)^2}{2!} = 0.9921$

c) $P(X = 4) = \dfrac{e^{-0.4}(0.4)^4}{4!} = 0.000715$

d) $P(X = 8) = \dfrac{e^{-0.4}(0.4)^8}{8!} = 1.09 \times 10^{-8}$

3-159. $P(X=0)=e^{-\lambda}=0.05$. Therefore, λ = –ln(0.05) = 2.996.
Consequently, E(X) = V(X) = 2.996.

3-160. a) Let X denote the number of calls in one hour. Then, X is a Poisson random variable with λ = 10.

$$P(X=5)=\frac{e^{-10}10^5}{5!}=0.0378.$$

b) $P(X\le 3)=e^{-10}+\frac{e^{-10}10}{1!}+\frac{e^{-10}10^2}{2!}+\frac{e^{-10}10^3}{3!}=0.0103$

c) Let Y denote the number of calls in two hours. Then, Y is a Poisson random variable with

E(Y) = 20. $P(Y=15)=\frac{e^{-20}20^{15}}{15!}=0.0516$

d) Let W denote the number of calls in 30 minutes. Then W is a Poisson random variable with

E(W) = 5. $P(W=5)=\frac{e^{-5}5^5}{5!}=0.1755$

3-161. λ=1, Poisson distribution. $f(x) = e^{-\lambda}\lambda^x/x!$
a) $P(X \ge 2) = 0.264$
b) In order that $P(X \ge 1) = 1 - P(X=0) = 1-e^{-\lambda}$ exceeds 0.95, we need λ = 3.
Therefore 3(16) = 48 cubic light years of space must be studied.

3-162. a) μ = 14.4, P(X = 0) = 6E10^{-7}
b) μ = 14.4/5 = 2.88, P(X = 0) = 0.056
c) μ = 14.4(7)(28.35)/225 = 12.7, $P(X \ge 1) = 0.999997$
d) $P(X \ge 28.8) = 1 - P(X \le 28) = 0.00046$. Unusual.

3-163. a) λ = 0.61 and $P(X \ge 1) = 0.4566$
b) μ = 0.61(5) = 3.05, P(X = 0) = 0.047

3-164.
a) Let X denote the number of flaws in one square meter of cloth. Then, X is a Poisson random variable with λ = 0.1.

$$P(X=2)=\frac{e^{-0.1}(0.1)^2}{2!}=0.0045$$

b) Let Y denote the number of flaws in 10 square meters of cloth. Then, Y is a Poisson random variable with E(Y) = 1.

$$P(Y=1)=\frac{e^{-1}1^1}{1!}=e^{-1}=0.3679$$

c) Let W denote the number of flaws in 20 square meters of cloth. Then, W is a Poisson random variable with E(W) = 2.

$$P(W=0)=e^{-2}=0.1353$$

d) $P(Y\ge 2)=1-P(Y\le 1)=1-P(Y=0)-P(Y=1)=1-e^{-1}-e^{-1}=0.2642$

3-165. a) $E(X)=0.2$ errors per test area

b) $P(X\le 2)=e^{-0.2}+\frac{e^{-0.2}0.2}{1!}+\frac{e^{-0.2}(0.2)^2}{2!}=0.9989$

99.89% of test areas

3-166. a) Let X denote the number of cracks in 5 miles of highway.
Then, X is a Poisson random variable with E(X) = 10.

$$P(X=0)=e^{-10}=4.54\times 10^{-5}$$

b) Let Y denote the number of cracks in a half mile of highway. Then, Y is a Poisson random variable with E(Y) = 1.

$$P(Y\ge 1)=1-P(Y=0)=1-e^{-1}=0.6321$$

c) The assumptions of a Poisson process require that the probability of an event is constant for all intervals. If the probability of a count depends on traffic load and the load varies, then the assumptions of a Poisson process are not valid. Separate Poisson random variables might be appropriate for the heavily and lightly loaded sections of the highway.

3-167. a) Let X denote the number of flaws in 10 square feet of plastic panel.
Then, X is a Poisson random variable with E(X) = 0.5.

$$P(X=0)=e^{-0.5}=0.6065$$

b) Let Y denote the number of cars with no flaws,

$$P(Y=10)=\binom{10}{10}(0.6065)^{10}(0.3935)^{0}=0.0067$$

c) Let W denote the number of cars with surface flaws. Because the number of flaws has a Poisson distribution, the occurrences of surface flaws in cars are independent events with constant probability. From part (a), the probability a car contains surface flaws is 1 − 0.6065 = 0.3935. Consequently, *W* has a binomial distribution with *n* = 10 and *p* = 0.3935

$$P(W=0)=\binom{10}{0}(0.3935)^{0}(0.6065)^{10}=0.0067$$

$$P(W=1)=\binom{10}{1}(0.3935)^{1}(0.6065)^{9}=0.0437$$

$$P(W\le 1)=0.0067+0.0437=0.0504$$

3-168. a) Let X denote the failures in 8 hours. Then, X has a Poisson distribution with E(X) = 0.16.

$$P(X=0)=e^{-0.16}=0.8521$$

b) Let Y denote the number of failure in 24 hours. Then, Y has a Poisson distribution with E(Y) = 0.48.

$$P(Y\ge 1)=1-P(Y=0)=1-e^{-48}=0.3812$$

3-169. a) $P(X\ge 2)=1-[P(X=0)+P(X=1)]=1-\left[\frac{e^{-0.25}0.25^{0}}{0!}+\frac{e^{-0.25}0.25^{1}}{1!}\right]=0.026$

b) $\lambda=0.25(5)=1.25$ per five days

$P(X=0)=e^{-1.25}=0.287$

c) $P(X\le 2)=P(X=0)+P(X=1)+P(X=2)$

$=e^{-1.25}+\frac{e^{-1.25}1.25}{1!}+\frac{e^{-1.25}1.25^{2}}{2!}=0.868$

3-170. a) $P(X=0)=e^{-1.5}=0.223$

b) $E(X)=1.5(10)=15$ per 10 minutes

$P(X\le 2)=P(X=0)+P(X=1)+P(X=2)$

$=e^{-15}+\frac{e^{-15}15}{1!}+\frac{e^{-15}15^{2}}{2!}=0.000039$

c) No, if a Poisson distribution is assumed, the intervals need not be consecutive.

3-171. a) Let *X* denote the number of cabs that pass your workplace in 10 minutes.

Then, X is a Poisson random variable with $\lambda T=5\frac{10}{60}=\frac{5}{6}$

$$\mathrm{P(X=0)}=\frac{e^{-5/6}(5/6)^{0}}{0!}=0.435$$

b) Let *Y* denote the number of cabs that pass your workplace in 20 minutes.

Then, *Y* is a Poisson random variable with $\lambda T=5\frac{20}{60}=\frac{5}{3}$

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \frac{e^{-5/3}(5/3)^0}{0!} = 0.811$$

c) Let λ^* be the mean number of cabs per hour and T = 1/6 hour. Find λ^* that satisfies the following condition:

$$P(X = 0) = \frac{e^{-\lambda^*/6}(\lambda^*/6)^0}{0!} = 0.1$$

$e^{-\lambda^*/6} = 0.1$ and $-\dfrac{\lambda^*}{6} = \ln(0.1)$

$\lambda^* = 13.816$

3-172. a) Let X denote the number of orders that arrive in 5 minutes.

Then, X is a Poisson random variable with $\lambda T = 12\dfrac{5}{60} = 1$.

$$P(X = 0) = \frac{e^{-1}(1)^0}{0!} = 0.368$$

b) $P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$

$$P(X \geq 3) = 1 - \left[\frac{e^{-1}(1)^0}{0!} + \frac{e^{-1}(1)^1}{1!} + \frac{e^{-1}(1)^2}{2!}\right] = 0.080$$

c) Let Y denote the number of orders arriving in the length of time T (in hours) that satisfies the condition. The mean of the random variable Y is $12T$ and T satisfies the following condition:

$$P(Y = 0) = \frac{e^{-12T}(12T)^0}{0!} = 0.001$$

$e^{-12T} = 0.001$ and $-12T = \ln(0.001)$

Therefore, T = 0.57565 hours = 34.54 minutes.

3-173. a) Let X denote the number of visits in a day.

Then, X is a Poisson random variable with $\lambda T = 1.8$

$$P(X > 5) = 1 - \sum_{n=0}^{5} P(X = n) = 1 - \sum_{n=0}^{5} \frac{e^{-1.8}(1.8)^n}{n!} = 0.010$$

b) Let Y denote the number of visits in a week.

Then, Y is a Poisson random variable with $\lambda T = 1.8(7) = 12.6$.

$$P(X < 5) = \sum_{n=0}^{4} P(X = n) = \sum_{n=0}^{4} \frac{e^{-12.6}(12.6)^n}{n!} = 0.005$$

c) Let Z denote the number of visits in T days that satisfies the given condition.
The mean of the random variable Z is $1.8T$.

$$P(Z \geq 1) = 1 - P(Z = 0) = 1 - \frac{e^{-1.8T}(1.8T)^0}{0!} = 0.99$$

$e^{-1.8T} = 0.01$ and $-1.8T = \ln(0.01)$. As a result, $T = 2.56$ days.

d) With T=1, determine λ such that $P(X > 5) = 1 - \sum_{n=0}^{5} P(X = n) = 1 - \sum_{n=0}^{5} \dfrac{e^{-\lambda}\lambda^n}{n!} = 0.1$

Solving the equation gives $\lambda = 3.15$

3-174. a) Let X denote the number of inclusions in cast iron with a volume of cubic millimeter.

Then, X is a Poisson random variable with $\lambda = 2.5$ and T = 1

$$\mathrm{P(X \ge 1) = 1 - P(X = 0)} = 1 - \frac{e^{-2.5}(2.5)^0}{0!} = 0.918$$

b) Let *Y* denote the number of inclusions in cast iron with a volume of 5.0 cubic millimeters. Then, *Y* is a Poisson random variable with $\lambda T = 2.5(5) = 12.5$

$$\mathrm{P(Y \ge 5)} = 1 - \sum_{n=0}^{4} \mathrm{P(X = n)} = 1 - \sum_{n=0}^{4} \frac{e^{-12.5}(12.5)^n}{n!} = 0.995$$

c) Let Z denote the number of inclusions in a volume of *V* cubic millimeters that satisfies the condition $\mathrm{P(Z \ge 1)} = 0.99$

The mean of the random variable Z is 2.5*V*.

$$\mathrm{P(Z \ge 1) = 1 - P(Z = 0)} = 1 - \frac{e^{-2.5V}(2.5V)^0}{0!} = 0.99.$$

As a result, *V*=1.84 cubic millimeters.

d) With *T* = 1, determine λ that satisfies

$$\mathrm{P(X \ge 1) = 1 - P(Z = 0)} = 1 - \frac{e^{-\lambda}(\lambda)^0}{0!} = 0.95$$

As a result, $\lambda = 3.00$ inclusions per cubic millimeter.

<u>Supplemental Exercises</u>

3-175. $$E(X) = \frac{1}{8}\left(\frac{1}{3}\right) + \frac{1}{4}\left(\frac{1}{3}\right) + \frac{3}{8}\left(\frac{1}{3}\right) = \frac{1}{4},$$

$$V(X) = \left(\frac{1}{8}\right)^2\left(\frac{1}{3}\right) + \left(\frac{1}{4}\right)^2\left(\frac{1}{3}\right) + \left(\frac{3}{8}\right)^2\left(\frac{1}{3}\right) - \left(\frac{1}{4}\right)^2 = 0.0104$$

3-176. a) $P(X = 1) = \binom{1000}{1} 0.001^1 (0.999)^{999} = 0.3681$

$$\text{b)} P(X \ge 1) = 1 - P(X = 0) = 1 - \binom{1000}{0} 0.001^0 (0.999)^{999} = 0.6319$$

$$\text{c)} P(X \le 2) = \binom{1000}{0} 0.001^0 (0.999)^{1000} + \binom{1000}{1} 0.001^1 (0.999)^{999} + \binom{1000}{2} 0.001^2 0.999^{998}$$
$$= 0.9198$$

$$\text{d)} E(X) = 1000(0.001) = 1$$
$$V(X) = 1000(0.001)(0.999) = 0.999$$

3-177. a) n = 50, p = 5/50 = 0.1, because E(X) = 5 = np

b) $P(X \le 2) = \binom{50}{0} 0.1^0 (0.9)^{50} + \binom{50}{1} 0.1^1 (0.9)^{49} + \binom{50}{2} 0.1^2 (0.9)^{48} = 0.112$

c) $P(X \ge 49) = \binom{50}{49} 0.1^{49} (0.9)^1 + \binom{50}{50} 0.1^{50} (0.9)^0 = 4.51 \times 10^{-48}$

3-178. a) Binomial distribution with p = 0.01, n = 12

b) P(X > 1) = 1 - P(X ≤ 1)= 1 - $\binom{12}{0} p^0 (1-p)^{12}$ - $\binom{12}{1} p^1 (1-p)^{14}$ = 0.0062

c) μ = E(X) = np =12(0.01) = 0.12

V(X) = np(1 - p) = 0.1188 σ = $\sqrt{V(X)}$ = 0.3447

3-179. a) $(0.5)^{12} = 0.000244$

b) $C_{12}^{6}(0.5)^6(0.5)^6$ = 0.2256

c) $C_5^{12}(0.5)^5(0.5)^7 + C_6^{12}(0.5)^6(0.5)^6 = 0.4189$

3-180. a) Binomial distribution with n =100, p = 0.01
b) P(X ≥ 1) = 0.634
c) P(X ≥ 2)= 0.264
d) μ = E(X) = np =100(0.01) = 1

V(X) = np(1 - p) = 0.99 and
σ= $\sqrt{V(X)}$ = 0.995

e) Let p_d= P(X≥2)= 0.264,
Y = number of messages that require two or more packets be resent.
Y is binomial distributed with n = 10, $p_m = p_d(1/10)$ = 0.0264
P(Y ≥ 1) = 0.235

3-181. Let X denote the number of mornings needed to obtain a green light.
Then X is a geometric random variable with p = 0.20.
a) P(X = 4) = $(1-0.2)^3 0.2$ = 0.1024
b) By independence, $(0.8)^{10}$ = 0.1074. (Also, P(X > 10) = 0.1074)

3-182. Let X denote the number of attempts needed to obtain a calibration that conforms to specifications.
Then, X is a geometric random variable with p = 0.6.
P(X ≤ 3) = P(X=1) + P(X=2) + P(X=3) = 0.6 + 0.4(0.6) + 0.4^2(0.6) = 0.936.

3-183. Let X denote the number of fills needed to detect three underweight packages.
Then, X is a negative binomial random variable with p = 0.001 and r = 3.
a) E(X) = 3/0.001 = 3000
b) V(X) = [3(0.999)/0.001^2] = 2997000. Therefore, σ_X = 1731.18

3-184. Geometric random variable with p = 0.1
a) f(x) = $(1-p)^{x-1}p = 0.9^{(x-1)}0.1$
b) P(X=5) = 0.9^4(0.1) = 0.0656
c) μ = E(X) = 1/p=10
d) P(X ≤ 10) = 0.651

3-185. a) E(X) = 6(0.5) = 3
P(X = 0) = 0.0498
b) P(X ≥ 3) = 0.5768
c) P(X ≤ x) ≥ 0.9, and by trial x = 5
d) σ^2 = λ = 6. Not appropriate.

3-186. Let X denote the number of totes in the sample that do not conform to purity requirements. Then, X has a hypergeometric distribution with N = 15, n = 3, and K = 2.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{2}{0}\binom{13}{3}}{\binom{15}{3}} = 1 - \frac{13!12!}{10!15!} = 0.3714$$

3-187. Let X denote the number of calls that are answered in 30 seconds or less.
Then, X is a binomial random variable with p = 0.75.

a) P(X = 9) = $\binom{10}{9}(0.75)^9(0.25)^1 = 0.1877$

b) P(X ≥ 16) = P(X = 16) +P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20)

$$= \binom{20}{16}(0.75)^{16}(0.25)^4 + \binom{20}{17}(0.75)^{17}(0.25)^3 + \binom{20}{18}(0.75)^{18}(0.25)^2$$

$$+ \binom{20}{19}(0.75)^{19}(0.25)^1 + \binom{20}{20}(0.75)^{20}(0.25)^0 = 0.4148$$

c) E(X) = 20(0.75) = 15

3-188. Let Y denote the number of calls needed to obtain an answer in less than 30 seconds.

a) $P(Y = 4) = (1-0.75)^3 0.75 = 0.25^3 0.75 = 0.0117$

b) E(Y) = 1/p = 1/0.75 = 4/3

3-189. Let W denote the number of calls needed to obtain two answers in less than 30 seconds.
Then, W has a negative binomial distribution with p = 0.75.

a) P(W = 6) = $\binom{5}{1}(0.25)^4(0.75)^2 = 0.0110$

b) E(W) = r/p = 2/0.75 = 8/3

3-190. a) Let X denote the number of messages sent in one hour.

$$P(X = 5) = \frac{e^{-5}5^5}{5!} = 0.1755$$

b) Let Y denote the number of messages sent in 1.5 hours.
Then, Y is a Poisson random variable with E(Y) = 7.5.

$$P(Y = 10) = \frac{e^{-7.5}(7.5)^{10}}{10!} = 0.0858$$

c) Let W denote the number of messages sent in one-half hour.
Then, W is a Poisson random variable with E(W) = 2.5

$$P(W < 2) = P(W = 0) + P(W = 1) = 0.2873$$

3-191. X is a negative binomial with r=4 and p=0.0001

$E(X) = r/p = 4/0.0001 = 40000$ requests

3-192. X ~ Poisson with E(X) = 0.01(100) = 1

$$P(Y \leq 3) = e^{-1} + \frac{e^{-1}(1)^1}{1!} + \frac{e^{-1}(1)^2}{2!} + \frac{e^{-1}(1)^3}{3!} = 0.9810$$

3-193. Let X denote the number of individuals that recover in one week. Assume the individuals are independent.
Then, X is a binomial random variable with n = 20 and p = 0.1.
P(X ≥ 4) = 1 − P(X ≤ 3) = 1 − 0.8670 = 0.1330.

3-194. a) P(X = 1) = 0, P(X = 2) = 0.0025, P(X = 3) = 0.01, P(X = 4) = 0.03, P(X = 5) = 0.065
P(X = 6) = 0.13, P(X = 7) = 0.18, P(X = 8) = 0.2225, P(X = 9) = 0.2, P(X = 10) = 0.16

b) P(X = 1) = 0.0025, P(X=1.5) = 0.01, P(X = 2) = 0.03, P(X = 2.5) = 0.065, P(X = 3) = 0.13
P(X = 3.5) = 0.18, P(X = 4) = 0.2225, P(X = 4.5) = 0.2, P(X = 5) = 0.16

3-195. Let X denote the number of assemblies needed to obtain 5 defectives.
Then, X is a negative binomial random variable with p = 0.01 and r=5.
a) E(X) = r/p = 500
b) V(X) = $5(0.99)/0.01^2$ = 49500 and σ = 222.49

3-196. Here n assemblies are checked. Let X denote the number of defective assemblies.
If $P(X \geq 1) \geq 0.95$, then $P(X = 0) \leq 0.05$. Now,

P(X = 0) = $\binom{n}{0}(0.01)^0(0.99)^n = 99^n$ and $0.99^n \leq 0.05$. Therefore,

$$n(\ln(0.99)) \leq \ln(0.05)$$

$$n \geq \frac{\ln(0.05)}{\ln(0.95)} = 298.07$$

Therefore, n = 299

3-197. Require f(1) + f(2) + f(3) + f(4) = 1. Therefore, c(1+2+3+4) = 1. Therefore, c = 0.1.

3-198. Let X denote the number of products that fail during the warranty period. Assume the units are independent. Then, X is a binomial random variable with n = 500 and p = 0.02.

a) P(X = 0) = $\binom{500}{0}(0.02)^0(0.98)^{500} = 4.1 \times 10^{-5}$

b) E(X) = 500(0.02) = 10
c) P(X >2) = 1 − P(X ≤ 2) = 0.9995

3-199.

$$f_X(0) = (0.1)(0.7) + (0.3)(0.3) = 0.16$$
$$f_X(1) = (0.1)(0.7) + (0.4)(0.3) = 0.19$$
$$f_X(2) = (0.2)(0.7) + (0.2)(0.3) = 0.20$$
$$f_X(3) = (0.4)(0.7) + (0.1)(0.3) = 0.31$$
$$f_X(4) = (0.2)(0.7) + (0)(0.3) = 0.14$$

3-200. a) P(X ≤ 3) = 0.2 + 0.4 = 0.6
b) P(X > 2.5) = 0.4 + 0.3 + 0.1 = 0.8
c) P(2.7 < X < 5.1) = 0.4 + 0.3 = 0.7
d) E(X) = 2(0.2) + 3(0.4) + 5(0.3) + 8(0.1) = 3.9
e) V(X) = $2^2(0.2) + 3^2(0.4) + 5^2(0.3) + 8^2(0.1) - (3.9)^2 = 3.09$

3-201.

x	2	5.7	6.5	8.5
f(x)	0.2	0.3	0.3	0.2

3-202. Let X and Y denote the number of bolts in the sample from supplier 1 and 2, respectively.
Then, X is a hypergeometric random variable with N = 100, n = 4, and K = 30.
Also, Y is a hypergeometric random variable with N = 100, n = 4, and K = 70.
a) P(X = 4 or Y=4) = P(X = 4) + P(Y = 4)

$$= \frac{\binom{30}{4}\binom{70}{0}}{\binom{100}{4}} + \frac{\binom{30}{0}\binom{70}{4}}{\binom{100}{4}}$$

$$= 0.2408$$

b) P[(X = 3 and Y = 1) or (Y = 3 and X = 1)]= $= \dfrac{\binom{30}{3}\binom{70}{1} + \binom{30}{1}\binom{70}{3}}{\binom{100}{4}} = 0.4913$

3-203. Let X denote the number of errors in a sector. Then, X is a Poisson random variable with E(X) = 0.32768.
a) $P(X > 1) = 1 - P(X \le 1) = 1 - e^{-0.32768} - e^{-0.32768}(0.32768) = 0.0433$
b) Let Y denote the number of sectors until an error is found.
Then, Y is a geometric random variable and $P = P(X \ge 1) = 1 - P(X = 0) = 1 - e^{-0.32768} = 0.2794$
E(Y) = 1/p = 3.58

3-204. Let X denote the number of orders placed in a week in a city of 800,000 people.
Then X is a Poisson random variable with E(X) = 0.25(8) = 2.
a) $P(X \ge 3) = 1 - P(X \le 2) = 1 - [e^{-2} + e^{-2}(2) + (e^{-2}2^2)/2!] = 1 - 0.6767 = 0.3233$.

b) Let Y denote the number of orders in 2 weeks. Then, Y is a Poisson random variable with E(Y) = 4, and
$P(Y > 2) =1- P(Y \le 2) = e^{-4} + (e^{-4}4^1)/1!+ (e^{-4}4^2)/2! =1 - [0.01832 + 0.07326 + 0.1465] = 0.7619$.

3-205. a) Hypergeometric random variable with N = 500, n = 5, and K = 125

$$f_X(0) = \frac{\binom{125}{0}\binom{375}{5}}{\binom{500}{5}} = \frac{6.0164E10}{2.5524E11} = 0.2357$$

$$f_X(1) = \frac{\binom{125}{1}\binom{375}{4}}{\binom{500}{5}} = \frac{125(8.10855E8)}{2.5525E11} = 0.3971$$

$$f_X(2) = \frac{\binom{125}{2}\binom{375}{3}}{\binom{500}{5}} = \frac{7750(8718875)}{2.5524E11} = 0.2647$$

$$f_X(3) = \frac{\binom{125}{3}\binom{375}{2}}{\binom{500}{5}} = \frac{317750(70125)}{2.5524E11} = 0.0873$$

$$f_X(4) = \frac{\binom{125}{4}\binom{375}{1}}{\binom{500}{5}} = \frac{9691375(375)}{2.5524E11} = 0.01424$$

$$f_X(5) = \frac{\binom{125}{5}\binom{375}{0}}{\binom{500}{5}} = \frac{2.3453E8}{2.5524E11} = 0.00092$$

b)

x	0	1	2	3	4	5
f(x)	0.0546	0.1866	0.2837	0.2528	0.1463	0.0574
	5	6	7	8	9	10
	0.0574	0.0155	0.0028	0.0003	0.0000	0.0000

3-206. Let X denote the number of totes in the sample that exceed the moisture content.
Then X is a binomial random variable with n = 30. We are to determine p.

If P(X ≥ 1) = 0.9, then P(X = 0) = 0.1. Then $\binom{30}{0}(p)^0(1-p)^{30} = 0.1$, and 30[ln(1−p)] = ln(0.1),

and p = 0.0739

3-207. Let T denote an interval of time in hours and let X denote the number of messages that arrive in time t.
Then, X is a Poisson random variable with E(X) = 10T.
Then, P(X=0) = 0.9 and $e^{-10T} = 0.9$, resulting in T = 0.0105 hours = 37.8 seconds

3-208. a) Let X denote the number of flaws in 50 panels.
Then, X is a Poisson random variable with E(X) = 50(0.02) = 1.
P(X = 0) = e^{-1} = 0.3679

b) Let Y denote the number of flaws in one panel.
P(Y ≥ 1) = 1 − P(Y = 0) = 1 − $e^{-0.02}$ = 0.0198
Let W denote the number of panels that need to be inspected before a flaw is found.
Then W is a geometric random variable with p = 0.0198.
E(W) = 1/0.0198 = 50.51 panels.

c) $P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-0.02} = 0.0198$
Let V denote the number of panels with 1 or more flaws.
Then V is a binomial random variable with n = 50 and p = 0.0198

$$P(V \leq 2) = \binom{50}{0}0.0198^0(.9802)^{50} + \binom{50}{1}0.0198^1(0.9802)^{49}$$
$$+\binom{50}{2}0.0198^2(0.9802)^{48} = 0.9234$$

3-209. a) Let *X* denote the number of cacti per 10,000 square meters.

Then, *X* is a Poisson random variable with E(X) = $280\frac{10{,}000}{10^6} = 2.8$

b) The unit is 10,000 square meters and *T* = 1. $\mathrm{P}(X = 0) = \frac{e^{-2.8}(2.8)^0}{0!} = 0.061$

c) Let *Y* denote the number of cacti in a region of area *T* (in units of 10,000 square meters).
The mean of the random variable *Y* is 2.8*T* and

$$\mathrm{P}(Y \geq 2) = 1 - [\mathrm{P}(Y = 0) + \mathrm{P}(Y = 1)] = 0.9$$

$$\frac{e^{-2.8T}(2.8T)^0}{0!}+\frac{e^{-2.8T}(2.8T)^1}{1!}=0.1$$

This can be solved in computer software to obtain $2.8T = 3.8897$
Therefore, T = 3.8897/2.8 = 1.39 (10,000 square meters) = 13,900 square meters.

3-210.

X = the number of sites with lesions in the sample
Then X has hypergeometric distribution with N = 50 and n = 8.
We need to find the minimum K that meets the condition $P(X \geq 1) \geq 0.95$
Equivalently, $1 - P(X = 0) \geq 0.95$ and $P(X = 0) \leq 0.05$ Therefore,

$$\frac{\binom{K}{0}\binom{50-K}{8-0}}{\binom{50}{8}} \leq 0.05$$

From trials of values for K, we have

$$\frac{\binom{15}{0}\binom{50-15}{8-0}}{\binom{50}{8}} < 0.05 < \frac{\binom{14}{0}\binom{50-14}{8-0}}{\binom{50}{8}}.$$

So, the minimum number of sites with lesions that satisfies the given condition is 15.

We also need to find the minimum K that meets the condition $P(X \geq 1) \geq 0.99$
Equivalently, $1 - P(X = 0) \geq 0.99$ and $P(X = 0) \leq 0.01$ Therefore,

$$\frac{\binom{K}{0}\binom{50-K}{8-0}}{\binom{50}{8}} \leq 0.01$$

From trials of values for K, we have

$$\frac{\binom{21}{0}\binom{50-21}{8-0}}{\binom{50}{8}} < 0.01 < \frac{\binom{20}{0}\binom{50-20}{8-0}}{\binom{50}{8}}.$$

So, the minimum number of sites with lesions that satisfies the given condition is 21.

Mind Expanding Exercises

3-211. The binomial distribution

$$P(X = x) = \frac{n!}{r!(n-r)!} p^x(1-p)^{n-x}$$

The probability of the event can be expressed as p = λ/n and the probability mass function can be written as

P(X = x) = $\frac{n!}{x!(n-x)!}[\lambda/n]^x[1-(\lambda/n)]^{n-x}$

P(X = x) $\frac{n\times(n-1)\times(n-2)\times(n-3)......\times(n-x+1)}{n^x}\frac{\lambda^x}{x!}(1-(\lambda/n))^{n-x}$

Now we can re-express as:

$[1-(\lambda/n)]^{n-x} = [1-(\lambda/n)]^n[1-(\lambda/n)]^{-x}$

In the limit as $n \to \infty$

$$\frac{n\times(n-1)\times(n-2)\times(n-3)......\times(n-x+1)}{n^x} \cong 1$$

As $n \to \infty$ the limit of $[1-(\lambda/n)]^{-x} \cong 1$

Also, we know that as $n \to \infty$

$$(1-\lambda/n)^n = e^{-\lambda}$$

Thus,

P(X = x) = $\frac{\lambda^x}{x!}e^{-\lambda}$

The distribution of the probability associated with this process is known as the Poisson distribution and we can express the probability mass function as

f(x) = $\frac{e^{-\lambda}\lambda^x}{x!}$

3-212. Show that $\sum_{i=1}^{\infty}(1-p)^{i-1}p = 1$ using an infinite sum.

To begin, $\sum_{i=1}^{\infty}(1-p)^{i-1}p = p\sum_{i=1}^{\infty}(1-p)^{i-1}$,

From the results for an infinite sum this equals

$$p\sum_{i=1}^{\infty}(1-p)^{i-1} = \frac{p}{1-(1-p)} = \frac{p}{p} = 1$$

3-213.

$$E(X) = [(a+(a+1)+\ldots+b](b-a+1)$$

$$= \left[\sum_{i=1}^{b} i - \sum_{i=1}^{a-1} i\right] \Big/ (b-a+1) = \left[\frac{b(b+1)}{2} - \frac{(a-1)a}{2}\right] \Big/ (b-a+1)$$

$$= \left[\frac{(b^2-a^2+b+a)}{2}\right] \Big/ (b-a+1) = \left[\frac{(b+a)(b-a+1)}{2}\right] \Big/ (b-a+1)$$

$$= \frac{(b+a)}{2}$$

$$V(X) = \frac{\sum_{i=a}^{b}\left[i - \frac{b+a}{2}\right]^2}{b+a-1} = \frac{\left[\sum_{i=a}^{b} i^2 - (b+a)\sum_{i=a}^{b} i + \frac{(b-a+1)(b+a)^2}{4}\right]}{b+a-1}$$

$$= \frac{\frac{b(b+1)(2b+1)}{6} - \frac{(a-1)a(2a-1)}{6} - (b+a)\left[\frac{b(b+1)-(a-1)a}{2}\right] + \frac{(b-a+1)(b+a)^2}{4}}{b-a+1}$$

$$= \frac{(b-a+1)^2-1}{12}$$

3-214. Let X denote a geometric random variable with parameter p. Let $q = 1 - p$.

$$E(X) = \sum_{x=1}^{\infty} x(1-p)^{x-1} p = p\sum_{x=1}^{\infty} xq^{x-1} = p\sum_{x=1}^{\infty} \frac{d}{dq} q^x$$

$$= p \cdot \frac{d}{dq} \sum_{x=1}^{\infty} q^x = p \cdot \frac{d}{dq}\left(\frac{q}{1-q}\right) = p\left(\frac{1(1-q)-q(-1)}{(1-q)^2}\right)$$

$$= p\left(\frac{1}{p^2}\right) = \frac{1}{p}$$

$$V(X) = \sum_{x=1}^{\infty} (x-\tfrac{1}{p})^2 (1-p)^{x-1} p = \sum_{x=1}^{\infty}\left(px^2 - 2x + \tfrac{1}{p}\right)(1-p)^{x-1}$$

$$= p\sum_{x=1}^{\infty} x^2 q^{x-1} - 2\sum_{x=1}^{\infty} xq^{x-1} + \tfrac{1}{p}\sum_{x=1}^{\infty} q^{x-1}$$

$$= p\sum_{x=1}^{\infty} x^2 q^{x-1} - \tfrac{2}{p^2} + \tfrac{1}{p^2}$$

$$= p\sum_{x=1}^{\infty} x^2 q^{x-1} - \tfrac{1}{p^2}$$

$$= p\,\tfrac{d}{dq}\left[q + 2q^2 + 3q^3 + \ldots\right] - \tfrac{1}{p^2}$$

$$= p\,\tfrac{d}{dq}\left[q(1+2q+3q^2+\ldots)\right] - \tfrac{1}{p^2}$$

$$= p\,\tfrac{d}{dq}\left[\tfrac{q}{(1-q)^2}\right] - \tfrac{1}{p^2} = 2pq(1-q)^{-3} + p(1-q)^{-2} - \tfrac{1}{p^2}$$

$$= \frac{\left[2(1-p)+p-1\right]}{p^2} = \frac{(1-p)}{p^2} = \frac{q}{p^2}$$

3-215.

Let X = number of passengers with a reserved seat who arrive for the flight,
n = number of seat reservations, p = probability that a ticketed passenger arrives for the flight.
a) In this part we determine n such that $P(X \geq 120) \geq 0.9$. By testing several values for n, the minimum value is n = 131.

b) In this part we determine n such that $P(X > 120) \leq 0.10$ which is equivalent to
$1 - P(X \leq 120) \leq 0.10$ or $0.90 \leq P(X \leq 120)$.
By testing several values for n, the solution is n = 123.

c) One possible answer follows. If the airline is most concerned with losing customers due to over-booking, they should only sell 123 tickets for this flight. The probability of over-booking is then at most 10%. If the airline is most concerned with having a full flight, they should sell 131 tickets for this flight. The chance the flight is full is then at least 90%. These calculations assume customers arrive independently and groups of people that arrive (or do not arrive) together for travel make the analysis more complicated.

3-216. Let X denote the number of nonconforming products in the sample.
Then, X is approximately binomial with p = 0.01 and n is to be determined.
If $P(X \geq 1) \geq 0.90$, then $P(X = 0) \leq 0.10$.
Now, $P(X = 0) = \binom{n}{0} p^0 (1-p)^n = (1-p)^n$. Consequently, $(1-p)^n \leq 0.10$, and
$n \leq \dfrac{\ln 0.10}{\ln(1-p)} = 229.11$. Therefore, n = 230 is required.

3-217. If the lot size is small, 10% of the lot might be insufficient to detect nonconforming product. For example, if the lot size is 10, then a sample of size one has a probability of only 0.2 of detecting a nonconforming product in a lot that is 20% nonconforming.

If the lot size is large, 10% of the lot might be a larger sample size than is practical or necessary. For example, if the lot size is 5000, then a sample of 500 is required. Furthermore, the binomial approximation to the hypergeometric distribution can be used to show the following. If 5% of the lot of size 5000 is nonconforming, then the probability of zero nonconforming products in the sample is approximately 7E-12. Using a sample of 100, the same probability is still only 0.0059. The sample of size 500 might be much larger than is needed.

3-218.

Let X denote the number of acceptable components. Then, X has a binomial distribution with p = 0.98 and n is to be determined such that $P(X \geq 100) \geq 0.95$

n	$P(X \geq 100)$
102	0.666
103	0.848
104	0.942
105	0.981

Therefore, 105 components are needed.

3-219. Let X denote the number of rolls produced.

Revenue at each demand				
	0	1000	2000	3000
$0 \leq x \leq 1000$	0.05x	0.3x	0.3x	0.3x
mean profit = 0.05x(0.3) + 0.3x(0.7) - 0.1x				
$1000 \leq x \leq 2000$	0.05x	0.3(1000) + 0.05(x-1000)	0.3x	0.3x
mean profit = 0.05x(0.3) + [0.3(1000) + 0.05(x-1000)](0.2) + 0.3x(0.5) - 0.1x				
$2000 \leq x \leq 3000$	0.05x	0.3(1000) + 0.05(x-1000)	0.3(2000) + 0.05(x-2000)	0.3x

mean profit = 0.05x(0.3) + [0.3(1000)+0.05(x-1000)](0.2) + [0.3(2000) + 0.05(x-2000)](0.3) + 0.3x(0.2) - 0.1x				
$3000 \le x$	0.05x	0.3(1000) + 0.05(x-1000)	0.3(2000) + 0.05(x-2000)	0.3(3000)+ 0.05(x-3000)
mean profit = 0.05x(0.3) + [0.3(1000)+0.05(x-1000)](0.2) + [0.3(2000)+0.05(x-2000)]0.3 + [0.3(3000)+0.05(x-3000)]0.2 - 0.1x				

	Profit	Max. profit
$0 \le x \le 1000$	0.125 x	$ 125 at x = 1000
$1000 \le x \le 2000$	0.075 x + 50	$ 200 at x = 2000
$2000 \le x \le 3000$	200	$200 at x = 3000
$3000 \le x$	-0.05 x + 350	$200 at x = 3000

The bakery can produce anywhere from 2000 to 3000 and earn the same profit.

CHAPTER 4

Section 4-2

4-1. a) $P(1 < X) = \int_1^{\infty} e^{-x}dx = (-e^{-x})\Big|_1^{\infty} = e^{-1} = 0.3679$

b) $P(1 < X < 2.5) = \int_1^{2.5} e^{-x}dx = (-e^{-x})\Big|_1^{2.5} = e^{-1} - e^{-2.5} = 0.2858$

c) $P(X = 3) = \int_3^{3} e^{-x}dx = 0$

d) $P(X < 4) = \int_0^{4} e^{-x}dx = (-e^{-x})\Big|_0^{4} = 1 - e^{-4} = 0.9817$

e) $P(3 \le X) = \int_3^{\infty} e^{-x}dx = (-e^{-x})\Big|_3^{\infty} = e^{-3} = 0.0498$

f) $P(x < X) = \int_x^{\infty} e^{-x}dx = (-e^{-x})\Big|_x^{\infty} = e^{-x} = 0.10$.

Then, x = –ln(0.10) = 2.3

g) $P(X \le x) = \int_0^{x} e^{-x}dx = (-e^{-x})\Big|_0^{x} = 1 - e^{-x} = 0.10$.

Then, x = –ln(0.9) = 0.1054

4-2. a) $P(X < 2) = \int_0^{2} \frac{3(8x - x^2)}{256}dx = (\frac{3x^2}{64} - \frac{x^3}{256})\Big|_0^{2} = (\frac{3}{16} - \frac{1}{32}) - 0 = 0.1563$

b) $P(X < 9) = \int_0^{8} \frac{3(8x - x^2)}{256}dx = (\frac{3x^2}{64} - \frac{x^3}{256})\Big|_0^{8} = (3 - 2) - 0 = 1$

c) $P(2 < X < 4) = \int_2^{4} \frac{3(8x - x^2)}{256}dx = (\frac{3x^2}{64} - \frac{x^3}{256})\Big|_2^{4} = (\frac{3}{4} - \frac{1}{4}) - (\frac{3}{16} - \frac{1}{32}) = 0.3438$

d) $P(X > 6) = \int_6^{8} \frac{3(8x - x^2)}{256}dx = (\frac{3x^2}{64} - \frac{x^3}{256})\Big|_6^{8} = (3 - 2) - (\frac{27}{16} - \frac{27}{32}) = 0.1563$

e) $P(X < x) = \int_0^{x} \frac{3(8u - u^2)}{256}du = (\frac{3u^2}{64} - \frac{u^3}{256})\Big|_0^{x} = (\frac{3x^2}{64} - \frac{x^3}{256}) - 0 = 0.95$

Then, $x^3 - 12x^2 + 243.2 = 0$, and $x = 6.9172$

4-3. a) $P(X<0)=\int_{-\pi/2}^{0} 0.5\cos x dx=(0.5\sin x)\Big|_{-\pi/2}^{0}=0-(-0.5)=0.5$

b) $P(X<-\pi/4)=\int_{-\pi/2}^{-\pi/4} 0.5\cos x dx=(0.5\sin x)\Big|_{-\pi/2}^{-\pi/4}=-0.3536-(-0.5)=0.1464$

c) $P(-\pi/4<X<\pi/4)=\int_{-\pi/4}^{\pi/4} 0.5\cos x dx=(0.5\sin x)\Big|_{-\pi/4}^{\pi/4}=0.3536-(-0.3536)=0.7072$

d) $P(X>-\pi/4)=\int_{-\pi/4}^{\pi/2} 0.5\cos x dx=(0.5\sin x)\Big|_{-\pi/4}^{\pi/2}=0.5-(-0.3536)=0.8536$

e) $P(X<x)=\int_{-\pi/2}^{x} 0.5\cos x dx=(0.5\sin x)\Big|_{-\pi/2}^{x}=(0.5\sin x)-(-0.5)=0.95$

Then, sin x = 0.9, and x = 1.1198 radians

4-4. a) $P(X<2)=\int_{1}^{2}\frac{2}{x^3}dx=(\frac{-1}{x^2})\Big|_{1}^{2}=(\frac{-1}{4})-(-1)=0.75$

b) $P(X>5)=\int_{5}^{\infty}\frac{2}{x^3}dx=(\frac{-1}{x^2})\Big|_{5}^{\infty}=0-(\frac{-1}{25})=0.04$

c) $P(4<X<8)=\int_{4}^{8}\frac{2}{x^3}dx=(\frac{-1}{x^2})\Big|_{4}^{8}=(\frac{-1}{64})-(\frac{-1}{16})=0.0469$

d) $P(X<4 or X>8)=1-P(4<X<8)$. From part (c), $P(4<X<8)$ = 0.0469. Therefore, $P(X<4 or X>8)$ = 1 – 0.0469 = 0.9531

e) $P(X<x)=\int_{1}^{x}\frac{2}{x^3}dx=(\frac{-1}{x^2})\Big|_{1}^{x}=(\frac{-1}{x^2})-(-1)=0.95$

Then, x^2 = 20, and x = 4.4721

4-5. a) $P(X<4)=\int_{3}^{4}\frac{x}{8}dx=\frac{x^2}{16}\Big|_{3}^{4}=\frac{4^2-3^2}{16}=0.4375$, because $f_X(x)=0$ for x < 3.

b) , $P(X>3.5)=\int_{3.5}^{5}\frac{x}{8}dx=\frac{x^2}{16}\Big|_{3.5}^{5}=\frac{5^2-3.5^2}{16}=0.7969$ because $f_X(x)=0$ for x > 5.

c) $P(4<X<5)=\int_{4}^{5}\frac{x}{8}dx=\frac{x^2}{16}\Big|_{4}^{5}=\frac{5^2-4^2}{16}=0.5625$

d) $P(X<4.5)=\int_{3}^{4.5}\frac{x}{8}dx=\frac{x^2}{16}\Big|_{3}^{4.5}=\frac{4.5^2-3^2}{16}=0.7031$

e) $P(X>4.5)+P(X<3.5)=\int_{4.5}^{5}\frac{x}{8}dx+\int_{3}^{3.5}\frac{x}{8}dx=\frac{x^2}{16}\Big|_{4.5}^{5}+\frac{x^2}{16}\Big|_{3}^{3.5}=\frac{5^2-4.5^2}{16}+\frac{3.5^2-3^2}{16}=0.5$.

4-6. a) $P(1 < X) = \int_4^\infty e^{-(x-4)}dx = -e^{-(x-4)}\Big|_4^\infty = 1$, because $f_X(x) = 0$ for x < 4. This can also be obtained from the fact that $f_X(x)$ is a probability density function for 4 < x.

b) $P(2 \le X \le 5) = \int_4^5 e^{-(x-4)}dx = -e^{-(x-4)}\Big|_4^5 = 1 - e^{-1} = 0.6321$

c) $P(5 < X) = 1 - P(X \le 5)$. From part b., $P(X \le 5) = 0.6321$. Therefore, $P(5 < X) = 0.3679$.

d) $P(8 < X < 12) = \int_8^{12} e^{-(x-4)}dx = -e^{-(x-4)}\Big|_8^{12} = e^{-4} - e^{-8} = 0.0180$

e) $P(X < x) = \int_4^x e^{-(x-4)}dx = -e^{-(x-4)}\Big|_4^x = 1 - e^{-(x-4)} = 0.90$.

Then, x = 4 – ln(0.10) = 6.303

4-7. a) $P(0 < X) = 0.5$, by symmetry.

b) $P(0.5 < X) = \int_{0.5}^1 1.5x^2dx = 0.5x^3\Big|_{0.5}^1 = 0.5 - 0.0625 = 0.4375$

c) $P(-0.5 \le X \le 0.5) = \int_{-0.5}^{0.5} 1.5x^2dx = 0.5x^3\Big|_{-0.5}^{0.5} = 0.125$

d) P(X < –2) = 0

e) P(X < 0 or X > –0.5) = 1

f) $P(x < X) = \int_x^1 1.5x^2dx = 0.5x^3\Big|_x^1 = 0.5 - 0.5x^3 = 0.05$

Then, x = 0.9655

4-8. a) $P(X > 3000) = \int_{3000}^\infty \frac{e^{\frac{-x}{1000}}}{1000}dx = -e^{\frac{-x}{1000}}\Big|_{3000}^\infty = e^{-3} = 0.05$

b) $P(1000 < X < 2000) = \int_{1000}^{2000} \frac{e^{\frac{-x}{1000}}}{1000}dx = -e^{\frac{-x}{1000}}\Big|_{1000}^{2000} = e^{-1} - e^{-2} = 0.233$

c) $P(X < 1000) = \int_0^{1000} \frac{e^{\frac{-x}{1000}}}{1000}dx = -e^{\frac{-x}{1000}}\Big|_0^{1000} = 1 - e^{-1} = 0.6321$

d) $P(X < x) = \int_0^x \frac{e^{\frac{-x}{1000}}}{1000}dx = -e^{\frac{-x}{1000}}\Big|_0^x = 1 - e^{-x/1000} = 0.10$.

Then, $e^{-x/1000} = 0.9$, and x = –1000 ln 0.9 = 105.36.

4-9. a) $P(X > 50) = \int_{50}^{50.25} 2.0dx = 2x\Big|_{50}^{50.25} = 0.5$

b) $P(X > x) = 0.90 = \int_{x}^{50.25} 2.0dx = 2x\Big|_{x}^{50.25} = 100.5 - 2x$

Then, 2x = 99.6 and x = 49.8.

4-10. a) $P(X < 74.8) = \int_{74.6}^{74.8} 1.25dx = 1.25x\Big|_{74.6}^{74.8} = 0.25$

b) P(X < 74.8 or X > 75.2) = P(X < 74.8) + P(X > 75.2) because the two events are mutually exclusive. The result is 0.25 + 0.25 = 0.50.

c) $P(74.7 < X < 75.3) = \int_{74.7}^{75.3} 1.25dx = 1.25x\Big|_{74.7}^{75.3} = 1.25(0.6) = 0.750$

4-11. a) P(X < 2.25 or X > 2.75) = P(X < 2.25) + P(X > 2.75) because the two events are mutually exclusive. Then, P(X < 2.25) = 0 and

P(X > 2.75) = $\int_{2.75}^{2.8} 2dx = 2(0.05) = 0.10$.

b) If the probability density function is centered at 2.55 meters, then $f_X(x) = 2$ for 2.3 < x < 2.8 and all rods will meet specifications.

4-12. a) $P(X < 90) = 0$ because the pdf is not defined in the range $(-\infty, 90)$.

b)

$$P(100 < X \le 200) = \int_{100}^{200} (-5.56\times10^{-4} + 5.56\times10^{-6}x)dx = (-5.56\times10^{-4}x + 2.78\times10^{-6}x^2)\Big|_{100}^{200}$$

$$= (-5.56\times10^{-4}\times200 + 2.78\times10^{-6}\times200^2) - (-5.56\times10^{-4}\times100 + 2.78\times10^{-6}\times100^2)$$

$$= 0.278$$

c)

$$P(X > 800) = \int_{800}^{1000} (4.44\times10^{-3} - 4.44\times10^{-6}x)dx = (4.44\times10^{-3}x - 2.22\times10^{-6}x^2)\Big|_{800}^{1000}$$

$$= (4.44\times10^{-3}\times10^3 - 2.22\times10^{-6}\times10^6) - (4.44\times10^{-3}\times800 - 2.22\times10^{-6}\times800^2)$$

$$= 0.0888 \cong 0.09$$

d) Find *a* such that $P(X > a) = 0.1$

$$P(X > a) = \int_{a}^{1000} (4.44\times10^{-3} - 4.44\times10^{-6}x)dx = (4.44\times10^{-3}x - 2.22\times10^{-6}x^2)\Big|_{a}^{1000} = 0.1$$

$$(4.44\times10^{-3}\times10^3 - 2.22\times10^{-6}\times10^6) - (4.44\times10^{-3}\times a - 2.22\times10^{-6}\times a^2) = 0.1$$

$$(2.22) - (4.44\times10^{-3}\times a - 2.22\times10^{-6}\times a^2) = 0.1$$

Then, $a \cong 787.76$

4-13. a) $P(0.5 < X) = \int_{0.5}^{2} 1 - 0.5x dx = x - 0.25x^2 \Big|_{0.5}^{2} = 0.562$

b) $P(a < X) = \int_{a}^{2} 1 - 0.5x dx = x - 0.25x^2 \Big|_{a}^{2} = 1 - (a - 0.25a^2) = 0.2 \Rightarrow a = 1.106$

c) $P(X = 0.22) = 0$

4-14. a) $P(X \le 40) = \int_{30}^{40} (0.0025x - 0.075)dx = (0.00125x^2 - 0.075x)\Big|_{30}^{40}$

$= (0.00125 \times 40^2 - 0.075 \times 40) - (0.00125 \times 30^2 - 0.075 \times 30) = 0.125$

b) $P(40 < X \le 60) = \int_{40}^{50} (0.0025x - 0.075)dx + \int_{50}^{60} (-0.0025x + 0.175)dx$

$= (0.00125x^2 - 0.075x)\Big|_{40}^{50} + (-0.00125x^2 + 0.175x)\Big|_{50}^{60}$

$= 0.375 + 0.375 = 0.75$

c) Find *a* such that $P(X > a) = 0.99$ (i.e. $P(X \le a) = 1 - 0.99 = 0.01$

$P(X \le a) = \int_{30}^{a} (0.0025x - 0.075)dx = (0.00125x^2 - 0.075x)\Big|_{30}^{a} = 0.01$

$= (0.00125 \times a^2 - 0.075 \times a) - (0.00125 \times 30^2 - 0.075 \times 30) = 0.01$

Then, $a = 32.828$

4-15. a) $P(X < 0.5) = \int_{0}^{0.5} 0.5\exp(-0.5x)dx = -\exp(-0.5x) \Big|_{0}^{0.5} = 0.221$

b) $P(X > 2) = \int_{2}^{\infty} 0.5\exp(-0.5x)dx = -\exp(-0.5x) \Big|_{2}^{\infty} = 0.368$

c) $P(X > a) = \int_{a}^{\infty} 0.5\exp(-0.5x)dx = -\exp(-0.5x) \Big|_{a}^{\infty} = \exp(-0.5a) = 0.05 \Rightarrow a = 5.991$

4-16. Because the integral $\int_{x_1}^{x_2} f(x)dx$ is not changed whether or not any of the endpoints x_1 and x_2 are included in the integral, all the probabilities listed are equal.

Section 4-3

4-17. a) P(X<2.8) = P(X ≤ 2.8) because X is a continuous random variable.
Then, P(X<2.8) =F(2.8)=0.2(2.8) = 0.56.
b) $P(X > 1.5) = 1 - P(X \le 1.5) = 1 - 0.2(1.5) = 0.7$
c) $P(X < -2) = F_X(-2) = 0$
d) $P(X > 6) = 1 - F_X(6) = 0$

4-18. a) $P(X < 1.8) = P(X \le 1.8) = F_X(1.8)$ because X is a continuous random variable. Then,
$$F_X(1.8) = 0.25(1.8) + 0.5 = 0.95$$
b) $P(X > -1.5) = 1 - P(X \le -1.5) = 1 - .125 = 0.875$
c) P(X < -2) = 0
d) $P(-1 < X < 1) = P(-1 < X \le 1) = F_X(1) - F_X(-1) = .75 - .25 = 0.50$

4-19. Now, $f(x) = e^{-x}$ for 0 < x and $F_X(x) = \int_0^x e^{-x} dx = -e^{-x}\Big|_0^x = 1 - e^{-x}$
for 0 < x. Then, $F_X(x) = \begin{cases} 0, x \le 0 \\ 1 - e^{-x}, x > 0 \end{cases}$

4-20. Now, $f(x) = \dfrac{3(8x - x^2)}{256}$ for 0 < x < 8 and
$$F_X(x) = \int_0^x \frac{3(8u - u^2)}{256} du = \left(\frac{3u^2}{64} - \frac{u^3}{256}\right)\Bigg|_0^x = \frac{3x^2}{64} - \frac{x^3}{256} \text{ for } 0 < x.$$
Then, $F_X(x) = \begin{cases} 0, \ x \le 0 \\ \dfrac{3x^2}{64} - \dfrac{x^3}{256}, 0 \le x < 8 \\ 1, \ x \ge 8 \end{cases}$

4-21. Now, $f(x) = 0.5\cos x$ for -π/2 < x < π/2 and
$$F_X(x) = \int_{-\pi/2}^{x} 0.5\cos u\, du = (0.5\sin x)\Big|_{-\pi/2}^{x} = 0.5\sin x + 0.5$$
Then, $F_X(x) = \begin{cases} 0, \ x \le -\pi/2 \\ 0.5\sin x + 0.5, -\pi/2 \le x < \pi/2 \\ 1, \ x \ge \pi/2 \end{cases}$

4-22. Now, $f(x) = \dfrac{2}{x^3}$ for x > 1 and

$$F_X(x) = \int_1^x \frac{2}{u^3}du = \left(\frac{-1}{u^2}\right)\Big|_1^x = \left(\frac{-1}{x^2}\right) + 1$$

Then, $F_X(x) = \begin{cases} 0, & x \le 1 \\ 1 - \dfrac{1}{x^2}, & x > 1 \end{cases}$

4-23. Now, $f(x) = x/8$ for 3 < x < 5 and $F_X(x) = \int_3^x \frac{u}{8}du = \frac{u^2}{16}\Big|_3^x = \frac{x^2 - 9}{16}$

for 0 < x. Then, $F_X(x) = \begin{cases} 0, & x < 3 \\ \dfrac{x^2 - 9}{16}, & 3 \le x < 5 \\ 1, & x \ge 5 \end{cases}$

4.24. Now, $f(x) = \dfrac{e^{-x/1000}}{1000}$ for 0 < x and

$$F_X(x) = \frac{1}{1000}\int_0^x e^{-x/1000}dy = -e^{-y/1000}\Big|_0^x = 1 - e^{-x/1000}$$ for 0 < x.

Then, $F_X(x) = \begin{cases} 0, & x \le 0 \\ 1 - e^{-x/1000}, & x > 0 \end{cases}$

P(X>3000) = 1- P(X ≤ 3000) = 1- F(3000) = $e^{-3000/1000}$ = 0.5

4-25. Now, f(x) = 2 for 2.3 < x < 2.8 and $F(x) = \int_{2.3}^x 2dy = 2x - 4.6$

for 2.3 < x < 2.8. Then,

$$F(x) = \begin{cases} 0, & x < 2.3 \\ 2x - 4.6, & 2.3 \le x < 2.8 \\ 1, & 2.8 \le x \end{cases}$$

$P(X > 2.7) = 1 - P(X \le 2.7) = 1 - F(2.7) = 1 - 0.8 = 0.2$ because X is a continuous random variable.

4-26. Now, $f(x) = \dfrac{e^{-x/10}}{10}$ for 0 < x and

$$F_X(x) = 1/10\int_0^x e^{-x/10}dx = -e^{-x/10}\Big|_0^x = 1 - e^{-x/10}$$

for 0 < x.

Then, $F_X(x) = \begin{cases} 0, & x \le 0 \\ 1 - e^{-x/10}, & x > 0 \end{cases}$

a) P(X<60) = F(60) =1 - e^{-6} = 1 - 0.002479 = 0.9975

b) $1/10\int_{15}^{30} e^{-x/10}dx = e^{-1.5} - e^{-3} = 0.173343$

c) $P(X_1>40) + P(X_1<40 \text{ and } X_2>40) = e^{-4} + (1 - e^{-4})\, e^{-4} = 0.0363$

d) P(15<X<30) = F(30)-F(15) = $e^{-1.5} - e^{-3} = 0.173343$

4-27. $F(x) = \int_0^x 0.5x dx = \frac{0.5x^2}{2}\Big|_0^x = 0.25x^2$ for 0 < x < 2. Then,

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.25x^2, & 0 \le x < 2 \\ 1, & 2 \le x \end{cases}$$

4-28. $f(x) = 2e^{-2x}, \; x > 0$

4-29. $f\ (x) = \begin{cases} 0.2, & 0 < x < 4 \\ 0.04, & 4 \le x < 9 \end{cases}$

4-30. $f_X(x) = \begin{cases} 0.25, & -2 < x < 1 \\ 0.5, & 1 \le x < 1.5 \end{cases}$

4-31.

$$f(x) = 1 - 0.5x, 0 < x < 2 \Rightarrow F(x) = \begin{cases} 0, x < 0 \\ \int_0^x 1 - 0.5x dx = x - 0.25x^2, 0 < x < 2 \\ 1, x > 2 \end{cases}$$

4-32. For $30 \le x < 50$,

$$P(X \le x) = \int_{30}^{x} (0.0025x - 0.075)dx = (0.00125x^2 - 0.075x)\Big|_{30}^{x}$$
$$= (0.00125x^2 - 0.075x) - (0.00125 \times 30^2 - 0.075 \times 30)$$
$$= 0.00125x^2 - 0.075x - 1.125$$

For $50 \le x < 70$,

$$P(X \le x) = F(50) + \int_{50}^{x} (-0.0025x + 0.175)dx$$
$$= 0.5 + (-0.00125x^2 + 0.175x)\Big|_{50}^{x}$$
$$= 0.5 + (-0.00125x^2 + 0.175x) - (-0.00125 \times 50^2 + 0.175 \times 50)$$
$$= 0.5 - 0.0125x^2 + 0.175x - 5.625 = -0.0125x^2 + 0.175x - 5.125$$

$$F(x) = \begin{cases} 0, x < 30 \\ 0.00125x^2 - 0.075x - 1.125, 30 \le x < 50 \\ -0.00125x^2 + 0.175x - 5.125, 50 \le x < 70 \\ 1, x \ge 70 \end{cases}$$

$$P(X \le 55) = F(55) = -0.00125 \times 55^2 + 0.175 \times 55 - 5.125 = 0.719$$

4-33. $f(x) = 0.5\exp(-0.5x) \Rightarrow F(x) = \begin{cases} 0, x < 0 \\ \int_{0}^{x} 0.5\exp(-0.5x)dx = 1 - \exp(-0.5x), x > 0 \end{cases}$

$$P(40 < X \le 60) = F(60) - F(40) = 2.06 \times 10^{-9}$$

4-34. For $100 \le x < 500$,

$$F(x) = \int_{100}^{x} (-5.56 \times 10^{-4} + 5.56 \times 10^{-6} x)dx = (-5.56 \times 10^{-4} x + 2.78 \times 10^{-6} x^2)\Big|_{100}^{x}$$
$$= (-5.56 \times 10^{-4} x + 2.78 \times 10^{-6} x^2) - (-5.56 \times 10^{-4} \times 100 + 2.78 \times 10^{-6} \times 100^2)$$

For $500 \le x < 1000$,

$$F(x) = F(500) + \int_{500}^{x} (4.44 \times 10^{-3} - 4.44 \times 10^{-6} x)dx$$
$$= 0.445 + (4.44 \times 10^{-3} x - 2.22 \times 10^{-6} x^2)\Big|_{500}^{x}$$
$$= 0.445 + (4.44 \times 10^{-3} x - 2.22 \times 10^{-6} x^2) - (4.44 \times 10^{-3} \times 500 - 2.22 \times 10^{-6} \times 500^2)$$
$$= 0.445 + (4.44 \times 10^{-3} x - 2.22 \times 10^{-6} x^2) - 1.665$$

$$F(x)=\begin{cases}0, x<100\\ 2.78\times10^{-6}x^2-5.56\times10^{-4}x+2.78\times10^{-2}, 100\le x<500\\ -2.22\times10^{-6}x^2+4.44\times10^{-3}x-1.22, 500\le x<1000\\ 1, x\ge 1000\end{cases}$$

<u>Section 4-4</u>

4-35. $E(X)=\int_0^4 0.25x\,dx=0.25\left.\frac{x^2}{2}\right|_0^4=2$

$$V(X)=\int_0^4 0.25(x-2)^2\,dx=0.25\left.\frac{(x-2)^3}{3}\right|_0^4=\frac{2}{3}+\frac{2}{3}=\frac{4}{3}$$

4-36. $E(X)=\int_0^4 0.125x^2dx=0.125\left.\frac{x^3}{3}\right|_0^4=2.6667$

$$V(X)=\int_0^4 0.125x(x-\tfrac{8}{3})^2\,dx=0.125\int_0^4(x^3-\tfrac{16}{3}x^2+\tfrac{64}{9}x)dx$$

$$=0.125\left.(\tfrac{x^4}{4}-\tfrac{16}{3}\tfrac{x^3}{3}+\tfrac{64}{9}\cdot\tfrac{1}{2}x^2)\right|_0^4=0.88889$$

4-37. $E(X)=\int_{-1}^1 1.5x^3dx=1.5\left.\frac{x^4}{4}\right|_{-1}^1=0$

$$V(X)=\int_{-1}^1 1.5x^3(x-0)^2\,dx=1.5\int_{-1}^1 x^4dx$$

$$=1.5\frac{x^5}{5}\Big|_{-1}^1=0.6$$

4-38. $E(X)=\int_3^5 x\frac{x}{8}dx=\left.\frac{x^3}{24}\right|_3^5=\frac{5^3-3^3}{24}=4.083$

$$V(X)=\int_3^5(x-4.083)^2\,\frac{x}{8}\,dx=\int_3^5\left(\frac{x^3}{8}-\frac{8.166x^2}{8}+\frac{16.6709x}{8}\right)dx$$

$$=\frac{1}{8}\left.\left(\frac{x^4}{4}-\frac{8.166x^3}{3}+\frac{16.6709x^2}{2}\right)\right|_3^5=0.3264$$

4-39. $E(X)=\int_0^8 x\frac{3(8x-x^2)}{256}dx=\left.(\frac{x^3}{32}-\frac{3x^4}{1024})\right|_0^8=(16-12)-0=4$

$$V(X)=\int_0^8 (x-4)^2\,\frac{3(8x-x^2)}{256}dx=\int_0^8\left(\frac{-3x^4}{256}+\frac{3x^3}{16}-\frac{15x^2}{16}+\frac{3x}{2}\right)dx$$

$$V(X)=\left(\frac{-3x^5}{1280}+\frac{3x^4}{64}-\frac{5x^3}{16}+\frac{3x^2}{4}\right)\Bigg|_0^8=\left(\frac{-384}{5}+192-160+48\right)=3.2$$

4-40. $E(X)=\int_1^\infty xe^{-x}dx$. Use integration by parts to obtain

$$E(X)=\int_0^\infty xe^{-x}dx=-e^{-x}(x+1)\Big|_0^\alpha=0-(-1)=1$$

$V(X)=\int_0^\infty (x-1)^2e^{-x}dx$ Use double integration by parts to obtain

$$V(X)=\int_0^\infty (x-1)^2e^{-x}dx=-(x-1)^2e^{-x}-2(x-1)e^{-x}-2e^{-x})\Big|_0^\alpha=1$$

4-41.

$$E(X)=\int_0^2 x(1-0.5x)dx=\frac{x^2}{2}-\frac{x^3}{6}\Bigg|_0^2=2-2\frac{1}{3}=-\frac{1}{3}$$

$$E(X^2)=\int_0^2 x^2(1-0.5x)dx=\frac{x^3}{3}-\frac{x^4}{8}\Bigg|_0^2=\frac{8}{3}-2=\frac{2}{3}$$

$$V(X)=E(X^2)-[E(X)]^2=\frac{2}{3}-\frac{1}{9}=\frac{5}{9}$$

4-42.

$$E(X)=\int_{30}^{50} x(0.0025x-0.075)dx=0.0025\frac{x^3}{3}-0.075\frac{x^2}{2}\Bigg|_{30}^{50}$$

$$+\int_{50}^{70} x(-0.0025x+0.175)dx=-0.0025\frac{x^3}{3}+0.175\frac{x^2}{2}\Bigg|_{50}^{70}==21\frac{2}{3}+28\frac{1}{3}=50$$

$$E(X^2)=\int_{30}^{50} x^2(0.0025x-0.075)dx=0.0025\frac{x^4}{4}-0.075\frac{x^3}{3}\Bigg|_{30}^{50}$$

$$+\int_{50}^{70} x^2(-0.0025x+0.175)dx=-0.0025\frac{x^4}{4}+0.175\frac{x^3}{3}\Bigg|_{50}^{70}=950+1616\frac{2}{3}=2566\frac{2}{3}$$

$$V(X)=E(X^2)-[E(X)]^2=2566\frac{2}{3}-2500=166\frac{1}{3}$$

4-43. Use integration by parts to obtain

$$E(X) = \int_0^\infty 0.5x\exp(-0.5x)dx = -x\exp(-0.5x)\Big|_0^{0.5} + \int_0^\infty \exp(-0.5x)dx$$
$$= -x\exp(-0.5x) - 2\exp(-0.5x)\ \Big|_0^{0.5} = 2$$

Use integration by parts two times to obtain

$$E(X^2) = \int_0^\infty 0.5x^2\exp(-0.5x)dx = 8$$
$$V(X) = E(X^2) - [E(X)]^2 = 8 - 2^2 = 4$$

4-44.
Probability distribution from exercise 4-12 used.

Here, a = 5.56E-6, b = -5.56E-4, c = -4.44E-6, d = 4.44E-3

$$E(X) = \int_{100}^{500} x(ax+b)dx = a\frac{x^3}{3} + b\frac{x^2}{2}\Big|_{100}^{500}$$
$$+ \int_{500}^{1000} x(cx+d)dx = c\frac{x^3}{3} + d\frac{x^2}{2}\Big|_{500}^{1000} = 163.0933 + 370 = 533.0933$$

$$E(X^2) = \int_{100}^{500} x^2(ax+b)dx = a\frac{x^4}{4} + b\frac{x^3}{3}\Big|_{100}^{500}$$
$$+ \int_{500}^{1000} x^2(cx+d)dx = d\frac{x^4}{4} + d\frac{x^3}{3}\Big|_{50}^{70} = 63754.67 + 254375 = 318129.7$$
$$V(X) = E(X^2) - [E(X)]^2 = 318129.7 - 533.0933^2 = 33941.16$$

4-45. $E(X) = \int_1^\infty x2x^{-3}dx = -2x^{-1}\Big|_1^\infty = 2$

4-46. a)

$$E(X) = \int_{1200}^{1210} x0.1dx = 0.05x^2\Big|_{1200}^{1210} = 1205$$
$$V(X) = \int_{1200}^{1210} (x-1205)^2 0.1dx = 0.1\frac{(x-1205)^3}{3}\Big|_{1200}^{1210} = 8.333$$
$$\sigma_x = \sqrt{V(X)} = 2.887$$

b) Clearly, centering the process at the center of the specifications results in the greatest proportion of cables within specifications.

$$P(1195 < X < 1205) = P(1200 < X < 1205) = \int_{1200}^{1205} 0.1dx = 0.1x\Big|_{1200}^{1205} = 0.5$$

4-47. a) $E(X)=\int_{100}^{120} x\frac{600}{x^2}dx = 600\ln x\Big|_{100}^{120} = 109.39$

$$V(X)=\int_{100}^{120}(x-109.39)^2\frac{600}{x^2}dx = 600\int_{100}^{120}1-\tfrac{2(109.39)}{x}+\tfrac{(109.39)^2}{x^2}dx$$

$$=600(x-218.78\ln x-109.39^2x^{-1})\Big|_{100}^{120}=33.19$$

b) Average cost per part = \$0.50*109.39 = \$54.70

4-48. a) $E(X)=\int_1^{70}xf(x)dx=\int_1^{70}\frac{70}{69x}dx=\frac{70}{69}\ln x\,|\begin{matrix}70\\1\end{matrix}=4.3101$

$$E(X^2)=\int_1^{70}x^2f(x)dx=\int_1^{70}\frac{70}{69}dx=70$$

Var(X) = $E(X^2)-(EX)^2=70-18.5770=51.4230$

b) 2.5(4.3101) = 10.7753

c) $P(X>50)=\int_{50}^{70}f(x)dx=0.0058$

4-49. a) $E(X)=\int_5^{\infty}x10e^{-10(x-5)}dx$.

Using integration by parts with $u = x$ and $dv=10e^{-10(x-5)}dx$, we obtain

$$E(X)=-xe^{-10(x-5)}\Big|_5^{\infty}+\int_5^{\infty}e^{-10(x-5)}dx=5-\frac{e^{-10(x-5)}}{10}\Bigg|_5^{\infty}=5.1$$

Now, $V(X)=\int_5^{\infty}(x-5.1)^2 10e^{-10(x-5)}dx$. Using the integration by parts with $u=(x-5.1)^2$ and

$dv=10e^{-10(x-5)}$, we obtain $V(X)=-(x-5.1)^2e^{-10(x-5)}\Big|_5^{\infty}+2\int_5^{\infty}(x-5.1)e^{-10(x-5)}dx$.

From the definition of E(X) the integral above is recognized to equal 0. Therefore,
$V(X)=(5-5.1)^2=0.01$.

b) $P(X>5.1)=\int_{5.1}^{\infty}10e^{-10(x-5)}dx=-e^{-10(x-5)}\Big|_{5.1}^{\infty}=e^{-10(5.1-5)}=0.3679$

Section 4-5

4-50. a) E(X) = (5.5 + 1.5)/2 = 3.5

$$V(X)=\frac{(5.5-1.5)^2}{12}=1.333,\ \ and\ \sigma_x=\sqrt{1.333}=1.155$$

b) $P(X<2.5)=\int_{1.5}^{2.5}0.25dx=0.25x\Big|_{1.5}^{2.5}=0.25$

c) $F(x)=\begin{cases} 0, & x<1.5 \\ 0.25x-0.375, & 1.5\le x<5.5 \\ 1, & 5.5\le x \end{cases}$

4-51. a) E(X) = (-1+1)/2 = 0

$$V(X)=\frac{(1-(-1))^2}{12}=1/3,\ and\ \sigma_x=0.577$$

b) $P(-x<X<x)=\int_{-x}^{x}\tfrac{1}{2}\,dt=0.5t\Big|_{-x}^{x}=0.5(2x)=x$

Therefore, x should equal 0.90.

c) $F(x)=\begin{cases} 0, & x<-1 \\ 0.5x+0.5, & -1\le x<1 \\ 1, & 1\le x \end{cases}$

4.52 a) $f(x)$ = 2.0 for 49.75 < x < 50.25.

E(X) = (50.25 + 49.75)/2 = 50.0

$$V(X)=\frac{(50.25-49.75)^2}{12}=0.0208,\ and\ \sigma_x=0.144.$$

b) $F(x)=\int_{49.75}^{x}2.0dy$ for 49.75 < x < 50.25. Therefore,

$$F(x)=\begin{cases} 0, & x<49.75 \\ 2x-99.5, & 49.75\le x<50.25 \\ 1, & 50.25\le x \end{cases}$$

c) $P(X<50.1)=F(50.1)=2(50.1)-99.5=0.7$

4-53. a) The distribution of X is $f(x)$ = 10 for 0.95 < x < 1.05. Now,

$$F_X(x)=\begin{cases} 0, & x<0.95 \\ 10x-9.5, & 0.95\le x<1.05 \\ 1, & 1.05\le x \end{cases}$$

b) $P(X>1.02)=1-P(X\le 1.02)=1-F_X(1.02)=0.3$

c) If P(X > x) = 0.90, then 1 – F(X) = 0.90 and F(X) = 0.10. Therefore, 10x - 9.5 = 0.10 and x = 0.96.

d) E(X) = (1.05 + 0.95)/2 = 1.00 and V(X) = $\dfrac{(1.05-0.95)^2}{12} = 0.00083$

4-54. $E(X) = \dfrac{(1.5+2.2)}{2} = 1.85\,\text{min}$

$V(X) = \dfrac{(2.2-1.5)^2}{12} = 0.0408\,\text{min}^{\,2}$

b) $P(X < 2) = \int_{1.5}^{2} \dfrac{1}{(2.2-1.5)}\,dx = \int_{1.5}^{2} (1/0.7)dx = (1/0.7)x\Big|_{1.5}^{2} = (1/0.7)(0.5) = 0.7143$

c.) $F(X) = \int_{1.5}^{x} \dfrac{1}{(2.2-1.5)}\,dy = \int_{1.5}^{x} (1/0.7)dy = (1/0.7)y\,\Big|_{1.5}^{x}$ for 1.5 < x < 2.2. Therefore,

$$F(x) = \begin{cases} 0, & x < 1.5 \\ (1/0.7)x - 2.14, & 1.5 \le x < 2.2 \\ 1, & 2.2 \le x \end{cases}$$

4-55. a) The distribution of X is *f*(*x*) = 100 for 0.2050 < x < 0.2150. Therefore,

$$F(x) = \begin{cases} 0, & x < 0.2050 \\ 100x - 20.50, & 0.2050 \le x < 0.2150 \\ 1, & 0.2150 \le x \end{cases}$$

b) $P(X > 0.2125) = 1 - F(0.2125) = 1 - [100(0.2125) - 20.50] = 0.25$

c) If P(X > x) = 0.10, then 1 – F(X) = 0.10 and F(X) = 0.90.
Therefore, 100x - 20.50 = 0.90 and x = 0.2140.

d) E(X) = (0.2050 + 0.2150)/2 = 0.2100 μm and
V(X) = $\dfrac{(0.2150-0.2050)^2}{12} = 8.33 \times 10^{-6}\,\mu\text{m}^2$

4-56. Let X denote the changed weight.
Var(X) = 4^2/12 and Stdev(X) = 1.1547

4-57. a) Let X be the time (in minutes) between arrival and 8:30 am.

$f(x) = \dfrac{1}{90}, \quad \textit{for} \quad 0 \le x \le 90$

Therefore, $F(x) = \dfrac{x}{90}, \quad \textit{for} \quad 0 \le x \le 90$

b) $E(X) = 45$, Var(X) = 90^2/12 = 675

c) The event is an arrival in the intervals 8:50-9:00 am or 9:20-9:30 am or 9:50-10:00 am so that the probability = 30/90 = 1/3

d) Similarly, the event is an arrival in the intervals 8:30-8:40 am or 9:00-9:10 am or 9:30-9:40 am so that the probability = 30/90 = 1/3

4-58. a) E(*X*) = (380 + 374)/2 = 377

$$V(X) = \frac{(380-374)^2}{12} = 3,\ and\ \sigma_x = 1.7321$$

b) Let *X* be the volume of a shampoo (milliliters)

$$P(X < 375) = \int_{374}^{375} \frac{1}{6} dx = \frac{1}{6} x \Big|_{374}^{375} = \frac{1}{6}(1) = 0.1667$$

c) The distribution of *X* is *f*(*x*) = 1/6 for $374 \le x \le 380$.

Now, $$F_X(x) = \begin{cases} 0, & x < 374 \\ (x-374)/6, & 374 \le x < 380 \\ 1, & 380 \le x \end{cases}$$

P(*X* > *x*) = 0.95, then 1 – F(*X*) = 0.95 and F(*X*) = 0.05.
Therefore, (*x* - 374)/6 = 0.05 and *x* = 374.3

d) Since E(*X*) = 377, then the mean extra cost = (377-375) x \$0.002 = \$0.004 per container.

4-59. (a) Let *X* be the arrival time (in minutes) after 9:00 A.M.

$$V(X) = \frac{(120-0)^2}{12} = 1200\ and\ \sigma_x = 34.64$$

b) We want to determine the probability the message arrives in any of the following intervals: 9:05-9:15 A.M. or 9:35-9:45 A.M. or 10:05-10:15 A.M. or 10:35-10:45 A.M.. The probability of this event is 40/120 = 1/3.

c) We want to determine the probability the message arrives in any of the following intervals: 9:15-9:30 A.M. or 9:45-10:00 A.M. or 10:15-10:30 A.M. or 10:45-11:00 A.M. The probability of this event is 60/120 = 1/2.

4-60. a) Let X denote the measured voltage.

Therefore, the probability mass function is $P(X = x) = \frac{1}{6}, \quad for \quad x = 247, \ldots, 253$

b) E(X)=250, Var(X)= $\frac{(253-247+1)^2-1}{12} = 4$

4-61. a) Yes. Time is uniformly distributed on any interval from 8:00 A.M. to 9:00 A.M.

b) $\mu = \frac{0+10}{2} = 5$

c) $P(X < 3) = \frac{3-0}{10-0} = 0.3$

4-62. Let X denote the kinetic energy of electron beams.

a) $\mu = E(X) = \frac{3+7}{2} = 5$

b) $\sigma^2 = V(X) = \frac{(7-3)^2}{12} \cong 1.33$

c) $P(X = 3.2) = 0$ because X has a continuous probability distribution.

d) $\mu = E(X) = \frac{3+b}{2} = 8$. Therefore, $b = 13$

e) $\sigma^2 = V(X) = \frac{(b-3)^2}{12} \cong 0.75$. Therefore, $b = 6$

<u>Section 4-6</u>

4-63. a) P(Z<1.32) = 0.90658
b) P(Z<3.0) = 0.99865
c) P(Z>1.45) = 1 – 0.92647 = 0.07353
d) P(Z > –2.15) = p(Z < 2.15) = 0.98422
e) P(–2.34 < Z < 1.76) = P(Z<1.76) – P(Z > 2.34) = 0.95116

4-64. a) P(–1 < Z < 1) = P(Z < 1) – P(Z > 1)
= 0.84134 – (1 – 0.84134) = 0.68268
b) P(–2 < Z < 2) = P(Z < 2) – [1 – P(Z < 2)] = 0.9545
c) P(–3 < Z < 3) = P(Z < 3) – [1 – P(Z < 3)] = 0.9973
d) P(Z > 3) = 1 – P(Z < 3) = 0.00135
e) P(0 < Z < 1) = P(Z < 1) – P(Z < 0) = 0.84134 – 0.5 = 0.34134

4-65. a) P(Z < 1.28) = 0.90
b) P(Z < 0) = 0.5
c) If P(Z > z) = 0.1, then P(Z < z) = 0.90 and z = 1.28
d) If P(Z > z) = 0.9, then P(Z < z) = 0.10 and z = –1.28
e) P(–1.24 < Z < z) = P(Z < z) – P(Z < –1.24) = P(Z < z) – 0.10749
Therefore, P(Z < z) = 0.8 + 0.10749 = 0.90749 and z = 1.33

4-66. a) Because of the symmetry of the normal distribution, the area in each tail of the distribution must equal 0.025. Therefore the value in Table III that corresponds to 0.975 is 1.96. Thus, z = 1.96.
b) Find the value in Table III corresponding to 0.995. z = 2.58
c) Find the value in Table III corresponding to 0.84. z = 1.0
d) Find the value in Table III corresponding to 0.99865. z = 3.0

4-67. a) P(X < 13) = P(Z < (13–10)/2) = P(Z < 1.5) = 0.93319
b) P(X > 9) = 1 – P(X < 9) = 1 – P(Z < (9–10)/2) = 1 – P(Z < –0.5) = 0.69146

c) P(6 < X < 14) = $P\left(\frac{6-10}{2} < Z < \frac{14-10}{2}\right)$ = P(–2 < Z < 2)

= P(Z < 2) –P(Z < – 2)]= 0.9545

d) P(2 < X < 4) = $P\left(\frac{2-10}{2} < Z < \frac{4-10}{2}\right)$

= P(–4 < Z < –3) = P(Z < –3) – P(Z < –4) = 0.00132

e) P(–2 < X < 8) = P(X < 8) – P(X < –2)
$= P\left(Z < \frac{8-10}{2}\right) - P\left(Z < \frac{-2-10}{2}\right)$ = P(Z < –1) – P(Z < –6) = 0.15866

4-68. a) P(X > x) = $P\left(Z > \frac{x-10}{2}\right)$ = 0.5. Therefore, $\frac{x-10}{2}$ = 0 and x = 10.

b) P(X > x) = $P\left(Z > \frac{x-10}{2}\right) = 1 - P\left(Z < \frac{x-10}{2}\right)$ = 0.95

Therefore, $P\left(Z < \frac{x-10}{2}\right)$ = 0.05 and $\frac{x-10}{2}$ = –1.64. Consequently, x = 6.72.

c) P(x < X < 10) = $P\left(\frac{x-10}{2} < Z < 0\right) = P(Z < 0) - P\left(Z < \frac{x-10}{2}\right)$

= 0.5 – $P\left(Z < \frac{x-10}{2}\right)$ = 0.2.

Therefore, $P\left(Z < \frac{x-10}{2}\right)$ = 0.3 and $\frac{x-10}{2}$ = –0.52. Consequently, x = 8.96.

d) P(10 – x < X < 10 + x) = P(–x/2 < Z < x/2)= 0.95. Therefore, x/2 = 1.96 and x = 3.92

e) P(10 – x < X < 10 + x) = P(–x/2 < Z < x/2) = 0.99. Therefore, x/2 = 2.58 and x = 5.16

4-69. a) P(X < 11) = $P\left(Z < \frac{11-5}{4}\right)$ = P(Z < 1.5) = 0.93319

b) P(X > 0) = $P\left(Z > \frac{0-5}{4}\right)$ = P(Z > –1.25) = 1 – P(Z < –1.25) = 0.89435

c) P(3 < X < 7) = $P\left(\frac{3-5}{4} < Z < \frac{7-5}{4}\right)$ = P(–0.5 < Z < 0.5)

= P(Z < 0.5) – P(Z < –0.5) = 0.38292

d) P(–2 < X < 9) = $P\left(\frac{-2-5}{4} < Z < \frac{9-5}{4}\right)$ = P(–1.75 < Z < 1)

= P(Z < 1) – P(Z < –1.75)] = 0.80128

e) P(2 < X < 8) = $P\left(\frac{2-5}{4} < Z < \frac{8-5}{4}\right)$ =P(–0.75 < Z < 0.75)

= P(Z < 0.75) – P(Z < –0.75) = 0.54674

4-70. a) P(X > x) = $P\left(Z > \frac{x-5}{4}\right)$ = 0.5. Therefore, x = 5.

b) P(X > x) = $P\left(Z > \frac{x-5}{4}\right)$ = 0.95. Therefore, $P\left(Z < \frac{x-5}{4}\right)$ = 0.05

Therefore, $\frac{x-5}{4}$ = –1.64, and x = –1.56.

c) P(x < X < 9) = $P\left(\frac{x-5}{4} < Z < 1\right)$ = 0.2.

Therefore, P(Z < 1) – P(Z < $\frac{x-5}{4}$)= 0.2 where P(Z < 1) = 0.84134.

Thus P(Z < $\frac{x-5}{4}$) = 0.64134. Consequently, $\frac{x-5}{4}$ = 0.36 and x = 6.44.

d) P(3 < X < x) = $P\left(\frac{3-5}{4} < Z < \frac{x-5}{4}\right)$ = 0.95.

Therefore, $P\left(Z < \frac{x-5}{4}\right)$ – P(Z < –0.5) = 0.95 and $P\left(Z < \frac{x-5}{4}\right)$ – 0.30854 = 0.95

Consequently,

$P\left(Z < \frac{x-5}{4}\right)$ = 1.25854.

Because a probability cannot be greater than one, there is no solution for x. In fact, P(3 < X) = P(–0.5 < Z) = 0.69146. Therefore, even if x is set to infinity the probability requested cannot equal 0.95.

e) *P*(– x < X – 5 < x) = *P*(5 – x < X < 5 + x) = $P\left(\frac{5-x-5}{4} < Z < \frac{5+x-5}{4}\right)$

= $P\left(\frac{-x}{4} < Z < \frac{x}{4}\right)$ = 0.99

Therefore, x/4 = 2.58 and x = 10.32.

4-71. a) P(X < 6250) = $P\left(Z < \frac{6250-6000}{100}\right)$ = P(Z < 2.5) = 0.99379

b) P(5800 < X < 5900) = $P\left(\frac{5800-6000}{100} < Z < \frac{5900-6000}{100}\right)$

= P(–2 < Z < –1) = P(Z <– 1) – P(Z < –2) = 0.13591

c) P(X > x) = $P\left(Z > \frac{x-6000}{100}\right)$ = 0.95. Therefore, $\frac{x-6000}{100}$ = –1.65 and x = 5835.

4-72. a) Let X denote the time.

X distributed N(260, 50^2)

$$P(X > 240) = 1 - P(X \le 240) = 1 - \Phi(\frac{240-260}{50}) = 1 - \Phi(-0.4) = 1 - \ 0.3446 = 0.6554$$

b) $\Phi^{-1}(0.25) \times 50 + 260 = 226.2755$

$\Phi^{-1}(0.75) \times 50 + 260 = 293.7245$

c) $\Phi^{-1}(0.05) \times 50 + 260 = 177.7550$

4-73. a) $1 - \Phi(2) = 0.0228$

b) Let X denote the time.

X ~ N(129, 14^2)

$$P(X < 100) = \Phi\left(\frac{100-129}{14}\right) = \Phi(-2.0714) = 0.01916$$

c) $\Phi^{-1}(0.95) \times 14 + 129 = 152.0280$

Here 95% of the surgeries will be finished within 152.028 minutes.

d) 199 >> 152.028 so the volume of such surgeries is very small (less than 5%).

4-74. Let X denote the cholesterol level.

$X \sim N(159.2, \sigma^2)$

a) $P(X < 200) = \Phi(\frac{200-159.2}{\sigma}) = 0.841$

$\frac{200-159.2}{\sigma} = \Phi^{-1}(0.841)$

$\sigma = \frac{200-159.2}{\Phi^{-1}(0.841)} = 40.8582$

b) $\Phi^{-1}(0.25) \times 40.8528 + 159.2 = 131.6452$

$\Phi^{-1}(0.75) \times 40.8528 + 159.2 = 186.7548$

c) $\Phi^{-1}(0.9) \times 40.8528 + 159.2 = 211.5550$

d) $\Phi(2) - \Phi(1) = 0.1359$

e) $1 - \Phi(2) = 0.0228$

f) $\Phi(-1) = 0.1587$

4-75. a) P(X > 0.62) = $P\left(Z > \frac{0.62-0.5}{0.05}\right)$ = P(Z > 2.4) = 1 – P(Z <2.4) = 0.0082

b) P(0.47 < X < 0.63) = $P\left(\frac{0.47-0.5}{0.05} < Z < \frac{0.63-0.5}{0.05}\right)$

= P(–0.6 < Z < 2.6) = P(Z < 2.6) – P(Z < –0.6) = 0.99534 – 0.27425 = 0.72109

c) P(X < x) = $P\left(Z < \frac{x-0.5}{0.05}\right)$ = 0.90

Therefore, $\frac{x-0.5}{0.05}$ = 1.28 and x = 0.564.

4-76. a) P(X < 12) = P(Z < $\frac{12-12.4}{0.1}$) = P(Z < –4) ≅ 0

b) P(X < 12.1) = $P\left(Z < \frac{12.1-12.4}{0.1}\right)$ = P(Z < –3) = 0.00135

and

P(X > 12.6) = $P\left(Z > \frac{12.6-12.4}{0.1}\right)$ = P(Z > 2) = 0.02275.

Therefore, the proportion of cans scrapped is 0.00135 + 0.02275 = 0.0241, or 2.41%

c) P(12.4 – x < X < 12.4 + x) = 0.99.

Therefore, $P\left(-\frac{x}{0.1} < Z < \frac{x}{0.1}\right) = 0.99$

Consequently, $P\left(Z < \frac{x}{0.1}\right) = 0.995$ and x = 0.1(2.58) = 0.258.

The limits are (12.142, 12.658).

4-77. a) If P(X > 12) = 0.999, then $P\left(Z > \frac{12-\mu}{0.1}\right) = 0.999$.

Therefore, $\frac{12-\mu}{0.1} = -3.09$ and $\mu = 12.309$.

b) If P(X > 12) = 0.999, then $P\left(Z > \frac{12-\mu}{0.05}\right) = 0.999$.

Therefore, $\frac{12-\mu}{0.05} = -3.09$ and $\mu = 12.1545$.

4-78. a) P(X > 0.5) = $P\left(Z > \frac{0.5-0.4}{0.05}\right)$ = P(Z > 2) = 1 – 0.97725 = 0.02275

b) P(0.4 < X < 0.5) = $P\left(\frac{0.4-0.4}{0.05} < Z < \frac{0.5-0.4}{0.05}\right)$

= P(0 < Z < 2) = P(Z < 2) – P(Z < 0) = 0.47725

c) P(X > x) = 0.90, then $P\left(Z > \frac{x-0.4}{0.05}\right) = 0.90$.

Therefore, $\frac{x-0.4}{0.05} = -1.28$ and x = 0.336.

4-79. a) P(X > 70) = $P\left(Z > \frac{70-60}{4}\right) = 1 - P(Z < 2.5) = 1 - 0.99379 = 0.00621$

b) P(X < 58) = $P\left(Z < \frac{58-60}{4}\right) = P(Z < -0.5) = 0.308538$

c) $1{,}000{,}000 \text{ bytes} * 8 \text{ bits/byte} = 8{,}000{,}000 \text{ bits}$

$$\frac{8{,}000{,}000 \text{ bits}}{60{,}000 \text{ bits/sec}} = 133.33 \text{ seconds}$$

4-80. Let X denote the height.

X ~ N(64, 2^2)

a) $P(58 < X < 70) = \Phi(\frac{70-64}{2}) - \Phi(\frac{58-64}{2}) = \Phi(3) - \Phi(-3) = 0.9973$

b) $\Phi^{-1}(0.25) \times 2 + 64 = 62.6510$

$\Phi^{-1}(0.75) \times 2 + 64 = 65.3490$

c) $\Phi^{-1}(0.05) \times 2 + 64 = 60.7103$

$\Phi^{-1}(0.95)\times 2+64=67.2897$

d) $[1-\Phi(\frac{68-64}{2})]^5=[1-\Phi(2)]^5=6.0942\times 10^{-9}$

4-81. Let X denote the height.

$X \sim N(1.41,\ 0.01^2)$

a) $P(X>1.42)=1-P(X\le 1.42)=1-\Phi(\frac{1.42-1.41}{0.01})=1-\Phi(1)=0.1587$

b) $\Phi^{-1}(0.05)\times 0.01+1.41=1.3936$

c) $P(1.39<X<1.43)=\Phi(\frac{1.43-1.41}{0.01})-\Phi(\frac{1.39-1.41}{0.01})=\Phi(2)-\Phi(-2)=0.9545$

4-82. Let X denote the demand for water daily.

$X \sim N(310,\ 45^2)$

a) $P(X>350)=1-P(X\le 350)=1-\Phi(\frac{350-310}{45})=1-\Phi(\frac{40}{45})=0.1870$

b) $\Phi^{-1}(0.99)\times 45+310=414.6857$

c) $\Phi^{-1}(0.05)\times 45+310=235.9816$

d) $X \sim N(\mu, 45^2)$

$P(X>350)=1-P(X\le 350)=1-\Phi(\frac{350-\mu}{45})=0.01$

$\Phi(\frac{350-\mu}{45})=0.99$

$\mu=350-\Phi^{-1}(0.99)\times 45=245.3143$ million gallons is the mean daily demand.

The mean daily demand per person = 245.3143/1.4 = 175.225 gallons.

4-83. a) $P(X<5000)=P\left(Z<\frac{5000-7000}{600}\right)=P(Z<-3.33)=0.00043$

b) $P(X>x)=0.95$. Therefore, $P\left(Z>\frac{x-7000}{600}\right)=0.95$ and $\frac{x-7000}{600}=-1.64$

Consequently, x = 6016

c) $P(X>7000)=P\left(Z>\frac{7000-7000}{600}\right)=P(Z>0)=0.5$

P(Three lasers operating after 7000 hours) = $(1/2)^3=1/8$

4-84. a) $P(X>0.0026)=P\left(Z>\frac{0.0026-0.002}{0.0004}\right)=P(Z>1.5)=1-P(Z<1.5)=0.06681$

b) $P(0.0014<X<0.0026)=P\left(\frac{0.0014-0.002}{0.0004}<Z<\frac{0.0026-0.002}{0.0004}\right)$

$=P(-1.5<Z<1.5)=0.86638$

c) P(0.0014 < X < 0.0026) = $P\left(\frac{0.0014-0.002}{\sigma} < Z < \frac{0.0026-0.002}{\sigma}\right)$

= $P\left(\frac{-0.0006}{\sigma} < Z < \frac{0.0006}{\sigma}\right)$

Therefore, $P\left(Z < \frac{0.0006}{\sigma}\right)$ = 0.9975. Therefore, $\frac{0.0006}{\sigma}$ = 2.81 and σ = 0.000214.

4-85. a) P(X > 13) = $P\left(Z > \frac{13-12}{0.5}\right)$ = P(Z > 2) = 0.02275

b) If P(X < 13) = 0.999, then $P\left(Z < \frac{13-12}{\sigma}\right)$ = 0.999

Therefore, 1/ σ = 3.09 and σ = 1/3.09 = 0.324

c) If P(X < 13) = 0.999, then $P\left(Z < \frac{13-\mu}{0.5}\right)$ = 0.999

Therefore, $\frac{13-\mu}{0.5}$ = 3.09 and μ = 11.455

4-86. a) Let X denote the measurement error, X ~ N(0, 0.5^2)

$$P(166.5 < 165.5 + X < 167.5) = P(1 < X < 2)$$

$$P(1 < X < 2) = \Phi\left(\frac{2}{0.5}\right) - \Phi\left(\frac{1}{0.5}\right) = \Phi(4) - \Phi(2) \approx 1 - 0.977 = 0.023$$

b) $P(166.5 < 165.5 + X) = P(1 < X)$

$$P(1 < X) = 1 - \Phi(1) = 1 - 0.841 = 0.159$$

4-87.

From the shape of the normal curve, the probability is maximized for an interval symmetric about the mean. Therefore *a* = 23.5 with probability = 0.1974. The standard deviation does not affect the choice of interval.

4-88. a) P(X > 9) = $P\left(Z > \frac{9-7.1}{1.5}\right)$ = P (Z > 1.2667) = 0.1026

b) $P(3 < X < 8) = P(X < 8) - P(X < 3) = P(Z < \frac{8-7.1}{1.5}) - P(Z < \frac{3-7.1}{1.5})$

= 0.7257 – 0.0031 = 0.7226.

c) P(X > x) = 0.05, then $\Phi^{-1}(0.95) \times 1.5 + 7.1 = 1.6449 \times 1.5 + 7.1 = 9.5673$

d) P(X > 9) = 0.01, then P(X < 9) = 1- 0.01 = 0.99

$P\left(Z < \frac{9-\mu}{1.5}\right) = 0.99$. Therefore, $\frac{9-\mu}{1.5} = 2.33$ and μ = 5.51.

4-89. a) P(X > 100) = $P\left(Z > \frac{100-50.9}{25}\right)$ = P (Z > 1.964) = 0.0248

b) P(X < 25) = $P\left(Z < \frac{25-50.9}{25}\right)$ = P (Z < -1.036) = 0.1501

c) P(X > x) = 0.05, then $\Phi^{-1}(0.95)\times 25 + 50.9 = 1.6449\times 25 + 50.9 = 92.0213$

4-90. a) P(X > 10) = $P\left(Z > \frac{10-4.6}{2.9}\right)$ = P (Z > 1.8621) = 0.0313

b) P(X > x) = 0.25, then $\Phi^{-1}(0.75)\times 2.9 + 4.6 = 0.6745\times 2.9 + 4.6 = 6.5560$

c) P(X < 0) = $P\left(Z < \frac{0-4.6}{2.9}\right)$ = P (Z < -1.5862) = 0.0563

The normal distribution is defined for all real numbers. In cases where the distribution is truncated (because wait times cannot be negative), the normal distribution may not be a good fit to the data.

4-91. $P(V > 1.5) = P(Z > \frac{1.5-\mu}{\sigma}) = 0.05 \Rightarrow \frac{1.5}{\sigma} = 1.645 \Rightarrow \sigma = 0.912$

4-92. Let X denote the value of the signal.

a) $P(1.45 \le X \le 1.55) = P\left(\frac{1.45-1.5}{0.02} \le Z \le \frac{1.55-1.5}{0.02}\right) = P(-2.5 \le Z \le 2.5) \cong 0.988$

b) $P(X > a) = 0.95$

$P\left(Z > \frac{a-1.5}{0.02}\right) = 0.95$ and $P\left(Z \le \frac{a-1.5}{0.02}\right) = 1-0.95 = 0.05$.

Then, $\frac{a-1.5}{0.02} = -1.645$ and $a = 1.4671$.

c)

$$P(X > \mu+2\sigma) = P\left(Z > \frac{\mu+2\sigma-\mu}{\sigma}\right) = P(Z > 2) = 1-P(Z \le 2) = 1-0.97725 = 0.02275$$

4-93. a) $P(Z > a) = 0.95 \Rightarrow a = -1.645$

b) $P(Z > 0) = 0.5$

c) $P(-1 < Z < 1) = F(1) - F(-1) = 0.841345 - 0.158655 = 0.683$

4-94.

Let X_1 and X_2 denote the right ventricle ejection fraction for PH subjects and control subjects, respectively.

a) Find a such that $P(X_1 > a) = 0.05$.

$P(X_1 > a) = P\left(Z > \frac{a-36}{12}\right) = 0.05$, then $P\left(Z \le \frac{a-36}{12}\right) = 1-0.05 = 0.95$

From the standard normal table, $\frac{a-36}{12} = 1.645$ and $a = 55.74$

b) $P(X_2 < 55.74) = P\left(Z_2 < \frac{55.74-56}{8}\right) = P(Z_2 < -0.0325) = 1-P(Z_2 < 0.0325) = 0.487$

c) Higher EF values are more likely to belong to the control subjects.

Section 4-7

4-95. a) E(X) = 200(0.4) = 80, V(X) = 200(0.4)(0.6) = 48 and $\sigma_X = \sqrt{48}$

Then, $P(X \le 70) \cong P\left(Z \le \frac{70.5-80}{\sqrt{48}}\right) = P(Z \le -1.37) = 0.0853$

b)

$$P(70 < X < 90) \cong P\left(\frac{70.5-80}{\sqrt{48}} < Z \le \frac{89.5-80}{\sqrt{48}}\right) = P(-1.37 < Z \le 1.37)$$
$$= 0.91466 - 0.08534 = 0.8293$$

c)

$$P(79.5 < X \le 80.5) \cong P\left(\frac{79.5-80}{\sqrt{48}} < Z \le \frac{80.5-80}{\sqrt{48}}\right) = P(-0.07217 < Z \le 0.07217)$$
$$= 0.0575$$

4-96. a) $P(X < 4) = \sum_{i=0}^{3} \frac{e^{-6}6^i}{i!} = 0.1512$

b) X is approximately $X \sim N(6,6)$

Then, $P(X < 4) \cong P\left(Z < \frac{4-6}{\sqrt{6}}\right) = P(Z < -0.82) = 0.206108$

If a continuity correction were used, the following result is obtained.

$$P(X < 4) = P(X \le 3) \cong P\left(Z \le \frac{3+0.5-6}{\sqrt{6}}\right) = P(Z \le -1.02) = 0.1539$$

c) $P(8 < X < 12) \cong P\left(\frac{8-6}{\sqrt{6}} < Z < \frac{12-6}{\sqrt{6}}\right) = P(0.82 < Z < 2.45) = 0.1990$

If a continuity correction were used, the following result is obtained.

$$P(8 < X < 12) = P(9 \le X \le 11) \cong P\left(\frac{9-0.5-6}{\sqrt{6}} \le Z \le \frac{11+0.5-6}{\sqrt{6}}\right)$$
$$\cong P(1.02 < Z < 2.25) = 0.1416$$

4-97. $Z = \frac{X-64}{\sqrt{64}} = \frac{X-64}{8}$ is approximately N(0,1).

a) $P(X > 72) = 1 - P(X \le 72) \cong 1 - P\left(Z \le \frac{72-64}{8}\right)$

$$= 1 - P(Z \le 1) = 1 - 0.8413 = 0.1587$$

If a continuity correction were used, the following result is obtained.

$$P(X > 72) = P(X \geq 73) \cong P\left(Z \geq \frac{73 - 0.5 - 64}{8}\right)$$

$$= P(Z \geq 1.06) = 1 - 0.855428 = 0.1446$$

b) 0.5

If a continuity correction were used, the following result is obtained.

$$P(X < 64) = P(X \leq 63) \cong P\left(Z \leq \frac{63 + 0.5 - 64}{8}\right) = P(Z \leq -0.06) = 0.4761$$

c)

$$P(60 < X \leq 68) = P(X \leq 68) - P(X \leq 60) = \Phi(\frac{68 - 64}{8}) - \Phi(\frac{60 - 64}{8})$$

$$= \Phi(0.5) - \Phi(-0.5) = 0.3829$$

If a continuity correction were used, the following result is obtained.

$$P(60 < X \leq 68) = P(61 \leq X \leq 68) \cong P\left(\frac{61 - 0.5 - 64}{8} < Z \leq \frac{68 + 0.5 - 64}{8}\right)$$

$$= P(-0.44 < Z \leq 0.56) = 0.3823$$

4-98. Let X denote the number of defective chips in the lot.

Then, E(X) = 1000(0.02) = 20, V(X) = 1000(0.02)(0.98) = 19.6.

a) $P(X > 25) \cong P\left(Z > \frac{25.5 - 20}{\sqrt{19.6}}\right) = P(Z > 1.24) = 1 - P(Z \leq 1.24) = 0.107$

b) $P(20 < X < 30) \cong P(20.5 < X < 29.5) = P(\frac{.5}{\sqrt{19.6}} < Z < \frac{9.5}{\sqrt{19.6}}) = P(0.11 < Z < 2.15)$

$$= 0.9842 - 0.5438 = 0.44$$

4-99. Let X denote the number of people with a disability in the sample.

X ~ Bin(1000, 0.193)

$Z = \frac{X - 1000 \times 0.193}{\sqrt{193(1 - 0.193)}} = \frac{X - 193}{12.4800}$ is approximately N(0,1).

a)

$$P(X > 200) = 1 - P(X \leq 200) = 1 - P(X \leq 200 + 0.5) = 1 - \Phi(\frac{200.5 - 193}{12.48}) = 1 - \Phi(0.6) = 0.2743$$

(b)

$$P(180 < X < 300) = P(181 \leq X \leq 299) = \Phi\left(\frac{299.5 - 193}{12.48}\right) - \Phi\left(\frac{180.5 - 193}{12.48}\right)$$

$$= \Phi(8.53) - \Phi(-1.00) = 0.8413$$

4-100. Let X denote the number of accounts in error in a month.

X ~ Bin(362000, 0.001)

a) E(X) = 362, Stdev(X) =19.0168

b) $Z = \dfrac{X - 362000 \times 0.001}{\sqrt{362(1-0.001)}} = \dfrac{X - 362}{19.0168}$ is approximately N(0,1).

$$P(X < 350) = P(X \le 349 + 0.5) = \Phi(\frac{349.5 - 362}{19.0168}) = \Phi(-0.6573) = 0.2555$$

c) $P(X \le v) = 0.95$

$$v = \Phi^{-1}(0.95) \times 19.0168 + 362 = 392.28$$

d)

$$P(X > 400) = 1 - P(X \le 400) = 1 - P(X \le 400 + 0.5) = 1 - \Phi(\frac{400.5 - 362}{19.0168}) = 1 - \Phi(2.0245) = 0.0215$$

Then the probability is $0.0215^2 = 4.6225 \times 10^{-4}$.

4-101. Let X denote the number of original components that fail during the useful life of the product. Then, X is a binomial random variable with p = 0.001 and n = 5000. Also, E(X) = 5000 (0.001) = 5 and V(X) = 5000(0.001)(0.999) = 4.995. With a continuity correction

$$P(X \ge 10) \cong P\left(Z \ge \frac{9.5 - 5}{\sqrt{4.995}} \right) = P(Z \ge 2.01) = 1 - P(Z < 2.01) = 1 - 0.978 = 0.022$$

4-102. Let X denote the number of errors on a web site.
Then, X is a binomial random variable with p = 0.05 and n = 100. Also, E(X) = 100 (0.05) = 5 and V(X) = 100(0.05)(0.95) = 4.75

$$P(X \ge 1) \cong P\left(Z \ge \frac{0.5 - 5}{\sqrt{4.75}} \right) = P(Z \ge -2.06) = 1 - P(Z < -2.06) = 1 - 0.0197 = 0.9803$$

4-103. Let X denote the number of particles in 10 cm^2 of dust. Then, X is a Poisson random variable with $\lambda T = 10(1000) = 10{,}000$. Also, E(X) = 10^4 and V(X) = 10^4

$$P(X > 10000) = 1 - P(X \le 10000) \cong 1 - P\left(Z \le \frac{10000 - 10000}{\sqrt{10000}} \right) \cong 1 - P(Z \le 0) \cong 0.5$$

With a continuity correction, the following result is obtained.

$$P(X > 10000) = P(X \ge 10001) \cong P\left(Z > \frac{10001 - 0.5 - 10000}{\sqrt{10000}} \right) \cong P(Z > 0) \cong 0.5$$

4-104. X is the number of minor errors on a test pattern of 1000 pages of text. Here X is a Poisson random variable with λ = 0.4 per page
a) The numbers of errors per page are random variables. The assumption that the occurrence of an event in a subinterval in a Poisson process is independent of events in other subintervals implies that the numbers of events in disjoint intervals are independent. The pages are disjoint intervals and the consequently the error counts per page are independent.

b) $P(X = 0) = \dfrac{e^{-0.4} 0.4^0}{0!} = 0.670$

$$P(X \ge 1) = 1 - P(X = 0) = 1 - 0.670 = 0.330$$

The mean number of pages with one or more errors is 1000(0.330) = 330 pages

c) Let Y be the number of pages with errors.

$$P(Y > 350) \cong P\left(Z \geq \frac{350.5-330}{\sqrt{1000(0.330)(0.670)}} \right) = P(Z \geq 1.38) = 1-P(Z < 1.38)$$
$$= 1-0.9162 = 0.0838$$

4-105. Let X denote the number of hits to a web site. Then, X is a Poisson random variable with a mean of 10,000 hits per day. Also, V(X) = 10,000.

a)

$$P(X > 20{,}000) = 1-P(X \leq 20{,}000) \cong 1-P\left(Z \leq \frac{20{,}000-10{,}000}{\sqrt{10{,}000}} \right)$$
$$= 1-P(Z \leq 100) \approx 1-1 = 0$$

If a continuity correction were used, the following result is obtained.

$$P(X > 20{,}000) = P(X \geq 20{,}001) \cong P\left(Z \geq \frac{20{,}001-0.5-10{,}000}{\sqrt{10{,}000}} \right)$$
$$= P(Z \geq 100.005) \approx 1-1 = 0$$

b) $P(X < 9{,}900) = P(X \leq 9{,}899) \cong P\left(Z \leq \frac{9{,}899-10{,}000}{\sqrt{10{,}000}} \right) = P(Z \leq -1.01) = 0.1562$

If a continuity correction were used, the following result is obtained.

$$P(X < 9{,}900) = P(X \leq 9{,}899) \cong P\left(Z < \frac{9{,}899+0.5-10{,}000}{\sqrt{10{,}000}} \right) = P(Z \geq -1.01) = 0.1562$$

c) If P($X > x$) = 0.01, then $P\left(Z > \frac{x-10{,}000}{\sqrt{10{,}000}} \right) = 0.01$.

Therefore, $\frac{x-10{,}000}{\sqrt{10{,}000}} = 2.33$ and $x = 10{,}233$

d) Let X denote the number of hits to a web site. Then, X is a Poisson random variable with a mean of 10,000 per day. Therefore, E(X) =10,000 and V(X) = 10,000

$$P(X > 10{,}200) \cong P\left(Z \geq \frac{10{,}200-10{,}000}{\sqrt{10{,}000}} \right) = P(Z \geq 2) = 1-P(Z < 2)$$
$$= 1-0.97725 = 0.02275$$

If a continuity correction were used, we obtain the following result

$$P(X > 10{,}200) \cong P\left(Z \geq \frac{10{,}200.5-10{,}000}{\sqrt{10{,}000}} \right) = P(Z \geq 2.005) = 1-P(Z < 2.005)$$

and this approximately equals the result without the continuity correction.

The expected number of days with more than 10,200 hits is (0.02275)*365 = 8.30 days per year.

e) Let Y denote the number of days per year with over 10,200 hits to a web site.
Then, Y is a binomial random variable with n = 365 and p = 0.02275.
E(Y) = 8.30 and V(Y) = 365(0.02275)(0.97725) = 8.28

$$P(Y > 15) \cong P\left(Z \geq \frac{15.5 - 8.30}{\sqrt{8.28}}\right) = P(Z \geq 2.56) = 1 - P(Z < 2.56)$$
$$= 1 - 0.9948 = 0.0052$$

4-106. Let X denotes the number of random sets that is more dispersed than the opteron.
Assume that X has a true mean = 0.5 x 1000 = 500 sets.

$$P(X \geq 750) \cong P\left(Z > \frac{750.5 - 1000(0.5)}{\sqrt{0.5(0.5)1000}}\right) = \left(Z > \frac{750.5 - 500)}{\sqrt{250}}\right)$$
$$= P(Z > 15.84) = 1 - P(Z \leq 15.84) \approx 0$$

4-107. With 10,500 asthma incidents in children in a 21-month period, then mean number of incidents per month is 10500/21 = 500. Let X denote a Poisson random variable with a mean of 500 per month.
Also, E(X) = λ = 500 = V(X).
a) Using a continuity correction, the following result is obtained.

$$P(X > 550) \cong P\left(Z \geq \frac{550 + 0.5 - 500}{\sqrt{500}}\right) = P(Z \geq 2.2584) = 1 - 0.9880 = 0.012$$

Without the continuity correction, the following result is obtained

$$P(X > 550) \cong P\left(Z \geq \frac{550 - 500}{\sqrt{500}}\right) = P(Z \geq 2.2361)$$
$$= 1 - P(Z < 2.2361) = 1 - 0.9873 = 0.0127$$

b) Using a continuity correction, the following result is obtained.

$$P(450 < X < 550) = P(451 \leq X \leq 549) \cong P\left(\frac{450.5 - 500}{\sqrt{500}} \leq Z \leq \frac{549.5 - 500}{\sqrt{500}}\right)$$
$$= P(Z \leq 2.2137) - P(Z \leq -2.2137) = 0.9866 - 0.0134 = 0.9732$$

Without the continuity correction, the following result is obtained

$$P(450 < X < 550) = P(X < 550) - P(X < 450) \cong P\left(Z \leq \frac{550 - 500}{\sqrt{500}}\right) - P\left(Z \leq \frac{450 - 500}{\sqrt{500}}\right)$$
$$= P(Z \leq 2.2361) - P(Z \leq -2.2361) = 0.9873 - 0.0127 = 0.9746$$

c) $P(X \leq x) = 0.95$

$$x = \Phi^{-1}(0.95) \times \sqrt{500} + 500 = 536.78$$

d) The Poisson distribution would not be appropriate because the rate of events should be constant for a Poisson distribution.

4-108. Approximate with a normal distribution.

$$np = 200 \times 0.3 = 60$$

$$\sqrt{np(1-p)} = \sqrt{42}$$

$$P(80 \le X) = P(80 - 0.5 \le X) \cong P\left(\frac{80 - 0.5 - 60}{\sqrt{42}} \le Z\right)$$

$$= 1 - P\left(Z \le \frac{80 - 0.5 - 60}{\sqrt{42}}\right) = 0.00135$$

If the medication has no effect, the probability that 80 or more patients experience relief of symptoms is small. Because 80 patients experienced relief of symptoms, we can conclude that the medication is most probably effective.

4-109. Approximate with a normal distribution.

$$np = 1500 \times 0.75 = 1125$$

$$\sqrt{np(1-p)} = 16.7705$$

a) $P(X \ge 1150) = P(Z > \frac{1150 - 0.5 - 1125}{16.7705}) = 0.072$

b) $P(1075 \le X \le 1175) = P(\frac{1075 - 0.5 - 1125}{16.7705} < Z < \frac{1175 + 0.5 - 1125}{16.7705}) = 0.997$

4-110. Let X denote the number of cabs that pass in a 10-hour day.

a) $\mu = E(X) = 5(10) = 50$ and $\sigma = \sqrt{5(10)} = \sqrt{50}$

b) $P(65 < X) = P(65 + 0.5 < X) \cong P\left(\frac{65 + 0.5 - 50}{\sqrt{50}} \le Z\right)$

$$= 1 - P\left(Z \le \frac{65 + 0.5 - 50}{\sqrt{50}}\right) = 0.014$$

c) $P(50 \le X \le 65) = P(50 \le X) - P(65 < X) \cong P\left(\frac{50 - 0.5 - 50}{\sqrt{50}} \le Z\right) - 0.02$

$$= 1 - P\left(Z \le \frac{50 - 0.5 - 50}{\sqrt{50}}\right) - 0.02 = 1 - 0.47 - 0.02 = 0.51$$

d) $P(100 \le X) = 0.95$

$$P(100 \le X) \cong P\left(\frac{100 - 0.5 - 10\lambda}{\sqrt{10\lambda}} \le Z\right) = 1 - P\left(Z < \frac{100 - 0.5 - 10\lambda}{\sqrt{10\lambda}}\right) = 0.95.$$

Therefore, $\frac{100 - 0.5 - 10\lambda}{\sqrt{10\lambda}} = -1.645$ and $\lambda \cong 11.74$.

4-111. a) $\mu = 2.5 \times 1000 = 2500, \sigma = \sqrt{2500} = 50$

b) $P(X < 2600) = P(X \le 2599) = P(Z < \frac{2599 + 0.5 - 2500}{50}) = 0.977$

c) $P(X > 2400) = P(X \geq 2401) = P(Z > \frac{2401 - 0.5 - 2500}{50}) = 0.977$

d) Consider the mean number of inclusions per cubic centimeter first.

$$P(X \leq 500) = P(Z < \frac{500 + 0.5 - \lambda}{\sqrt{\lambda}}) = 0.9 \Rightarrow \frac{500 + 0.5 - \lambda}{\sqrt{\lambda}} = 1.28 \Rightarrow \lambda = 472.67$$

Therefore, the mean number of inclusions per cubic millimeter is 472.67/1000 = 0.473

Section 4-8

4-112. a) $P(X \leq 0) = \int_0^0 \lambda e^{-\lambda x} dx = 0$

b) $P(X \geq 2) = \int_2^\infty 2e^{-2x} dx = -e^{-2x}\Big|_2^\infty = e^{-4} = 0.0183$

c) $P(X \leq 1) = \int_0^1 2e^{-2x} dx = -e^{-2x}\Big|_0^1 = 1 - e^{-2} = 0.8647$

d) $P(1 < X < 2) = \int_1^2 2e^{-2x} dx = -e^{-2x}\Big|_1^2 = e^{-2} - e^{-4} = 0.1170$

e) $P(X \leq x) = \int_0^x 2e^{-2t} dt = -e^{-2t}\Big|_0^x = 1 - e^{-2x} = 0.05$ and x = 0.0256

4-113. If E(X) = 10, then $\lambda = 0.1$.

a) $P(X > 10) = \int_{10}^\infty 0.1e^{-0.1x} dx = -e^{-0.1x}\Big|_{10}^\infty = e^{-1} = 0.3679$

b) $P(X > 20) = -e^{-0.1x}\Big|_{20}^\infty = e^{-2} = 0.1353$

c) $P(X < 30) = -e^{-0.1x}\Big|_0^{30} = 1 - e^{-3} = 0.9502$

d) $P(X < x) = \int_0^x 0.1e^{-0.1t} dt = -e^{-0.1t}\Big|_0^x = 1 - e^{-0.1x} = 0.95$ and x = 29.96.

4-114. a) $P(X < 5) = 0.3935$

b) $P(X < 15 \mid X > 10) = \frac{P(X < 15, \quad X > 10)}{P(X > 10)} = \frac{P(X < 15) - P(X < 10)}{1 - P(X < 10)} = \frac{0.1447}{0.3679} = 0.3933$

c) They are the same.

4-115. Let X denote the time until the first count. Then, X is an exponential random variable with $\lambda = 2$ counts per minute.

a) $P(X > 0.5) = \int_{0.5}^{\infty} 2e^{-2x} dx = -e^{-2x}\Big|_{0.5}^{\infty} = e^{-1} = 0.3679$

b) $P(X < \frac{10}{60}) = \int_{0}^{1/6} 2e^{-2x} dx = -e^{-2x}\Big|_{0}^{1/6} = 1 - e^{-1/3} = 0.2835$

c) $P(1 < X < 2) = -e^{-2x}\Big|_{1}^{2} = e^{-2} - e^{-4} = 0.1170$

4-116. a) E(X) = 1/λ = 1/3 = 0.333 minutes
b) V(X) = $1/\lambda^2 = 1/3^2 = 0.111$, σ = 0.3333
c) $P(X < x) = \int_{0}^{x} 3e^{-3t} dt = -e^{-3t}\Big|_{0}^{x} = 1 - e^{-3x} = 0.95$, x = 0.9986

4-117. Let X denote the time until the first call. Then, X is exponential and $\lambda = \frac{1}{E(X)} = \frac{1}{15}$ calls/minute.

a) $P(X > 30) = \int_{30}^{\infty} \frac{1}{15} e^{-\frac{x}{15}} dx = -e^{-\frac{x}{15}}\Big|_{30}^{\infty} = e^{-2} = 0.1353$

b) The probability of at least one call in a 10-minute interval equals one minus the probability of zero calls in a 10-minute interval and that is P(X > 10).

$P(X > 10) = -e^{-\frac{x}{15}}\Big|_{10}^{\infty} = e^{-2/3} = 0.5134$.

Therefore, the answer is 1- 0.5134 = 0.4866. Alternatively, the requested probability is equal to P(X < 10) = 0.4866.

c) $P(5 < X < 10) = -e^{-\frac{x}{15}}\Big|_{5}^{10} = e^{-1/3} - e^{-2/3} = 0.2031$

d) P(X < x) = 0.90 and $P(X < x) = -e^{-\frac{t}{15}}\Big|_{0}^{x} = 1 - e^{-x/15} = 0.90$. Therefore, x = 34.54 minutes.

4-118. Let X be the life of regulator. Then, X is an exponential random variable with $\lambda = 1/E(X) = 1/6$
a) Because the Poisson process from which the exponential distribution is derived is memoryless, this probability is

P(X < 6) = $\int_{0}^{6} \frac{1}{6} e^{-x/6} dx = -e^{-x/6}\Big|_{0}^{6} = 1 - e^{-1} = 0.6321$

b) Because the failure times are memoryless, the mean time until the next failure is E(X) = 6 years.

4-119. Let X denote the time to failure (in hours) of fans in a personal computer. Then, X is an exponential random variable and $\lambda = 1/E(X) = 0.0003$.

a) P(X > 10,000) = $\int_{10,000}^{\infty} 0.0003 e^{-x0.0003} dx = -e^{-x0.0003}\Big|_{10,000}^{\infty} = e^{-3} = 0.0498$

b) P(X < 7,000) = $\int_0^{7,000} 0.0003 e^{-x0.0003} dx = -e^{-x0.0003}\Big|_0^{7,000} = 1 - e^{-2.1} = 0.8775$

4-120. Let X denote the time until a message is received. Then, X is an exponential random variable and $\lambda = 1/E(X) = 1/2$

a) P(X > 2) = $\int_2^{\infty} \tfrac{1}{2} e^{-x/2} dx = -e^{-x/2}\Big|_2^{\infty} = e^{-1} = 0.3679$

b) The same as part a).

c) E(X) = 2 hours.

4-121. Let X denote the time until the arrival of a taxi. Then, X is an exponential random variable with $\lambda = 1/E(X) = 0.1$ arrivals/ minute.

a) P(X > 60) = $\int_{60}^{\infty} 0.1 e^{-0.1x} dx = -e^{-0.1x}\Big|_{60}^{\infty} = e^{-6} = 0.0025$

b) P(X < 10) = $\int_0^{10} 0.1 e^{-0.1x} dx = -e^{-0.1x}\Big|_0^{10} = 1 - e^{-1} = 0.6321$

c) P(X > x) = $\int_x^{\infty} 0.1 e^{-0.1t} dt = -e^{-0.1t}\Big|_x^{\infty} = e^{-0.1x} = 0.1$ and x = 23.03 minutes.

d) P(X < x) = 0.9 implies that P(X > x) = 0.1. Therefore, this answer is the same as part c).

e) P(X < x) = $-e^{-0.1t}\Big|_0^{x} = 1 - e^{-0.1x} = 0.5$ and x = 6.93 minutes.

4-122. a) 1/2.3 = 0.4348 year

b) E(X) = 2.3(0.25) = 0.575, P(X=0) = exp(-0.575) = 0.5627

c) Let T denote the time between sightings.
Here T has an exponential distribution with mean 0.4348.
$P(X > 0.5) = 1 - P(X \le 0.5) = 0.3167$

d) E(X) = 2.3(3) = 6.9, P(X = 0) = exp(-6.9) = 0.001

4.123. Let X denote the number of insect fragments per gram. Then λ = 14.4/225

a) 225/14.4 = 15.625

b) P(X = 0) = exp[-14.4(28.55)/225] = 0.1629

c) $(0.1629)^7 = 3 \times 10^{-6}$

4-124. Let X denote the distance between major cracks. Then, X is an exponential random variable with $\lambda = 1/E(X) = 0.2$ cracks/mile.

a) P(X > 10) = $\int_{10}^{\infty} 0.2e^{-0.2x}dx = -e^{-0.2x}\Big|_{10}^{\infty} = e^{-2} = 0.1353$

b) Let Y denote the number of cracks in 10 miles of highway. Because the distance between cracks is exponential, Y is a Poisson random variable with E(Y) = 10(0.2) = 2 cracks per 10 miles.

P(Y = 2) = $\dfrac{e^{-2}2^2}{2!} = 0.2707$

c) $\sigma_X = 1/\lambda = 5$ miles.

d) $P(12 < X < 15) = \int_{12}^{15} 0.2e^{-0.2x}dx = -e^{-0.2x}\Big|_{12}^{15} = e^{-2.4} - e^{-3} = 0.0409$

e) P(X > 5) = $-e^{-0.2x}\Big|_{5}^{\infty} = e^{-1} = 0.3679$.

By independence of the intervals in a Poisson process, the answer is $0.3679^2 = 0.1353$.
Alternatively, the answer is P(X > 10) = $e^{-2} = 0.1353$. The probability does depend on whether or not the lengths of highway are consecutive.

f) By the memoryless property, this answer is P(X > 10) = 0.1353 from part e).

4-125. Let X denote the lifetime of an assembly. Then, X is an exponential random variable with $\lambda = 1/E(X) = 1/400$ failures per hour.

a) P(X < 100) = $\int_{0}^{100} \frac{1}{400}e^{-x/400}dx = -e^{-x/400}\Big|_{0}^{100} = 1 - e^{-0.25} = 0.2212$

b) P(X > 500) = $-e^{-x/400}\Big|_{500}^{\infty} = e^{-5/4} = 0.2865$

c) From the memoryless property of the exponential, this answer is the same as part a.,
P(X < 100) = 0.2212.

d) Let U denote the number of assemblies out of 10 that fail before 100 hours. By the memoryless property of a Poisson process, U has a binomial distribution with n = 10 and
p = 0.2212 from part a). Then,

$P(U \geq 1) = 1 - P(U = 0) = 1 - \binom{10}{0}0.2212^0(1-0.2212)^{10} = 0.9179$

e) Let V denote the number of assemblies out of 10 that fail before 800 hours. Then, V is a binomial random variable with n = 10 and p = P(X < 800), where X denotes the lifetime of an assembly.

Now, P(X < 800) = $\int_{0}^{800} \frac{1}{400}e^{-x/400}dx = -e^{-x/400}\Big|_{0}^{800} = 1 - e^{-2} = 0.8647$.

Therefore, P(V = 10) = $\binom{10}{10}0.8647^{10}(1-0.8647)^0 = 0.2337$.

4-126. Let Y denote the number of arrivals in one hour. If the time between arrivals is exponential, then the count of arrivals is a Poisson random variable and $\lambda = 1$ arrival per hour.

a) P(Y > 3) = $1-P(Y\le 3)=1-\left[\frac{e^{-1}1^0}{0!}+\frac{e^{-1}1^1}{1!}+\frac{e^{-1}1^2}{2!}+\frac{e^{-1}1^3}{3!}\right]=0.01899$

b) From part a), P(Y > 3) = 0.01899. Let W denote the number of one-hour intervals out of 30 that contain more than 3 arrivals. By the memoryless property of a Poisson process, W is a binomial random variable with n = 30 and p = 0.01899.

P(W = 0) = $\binom{30}{0}0.01899^0(1-0.01899)^{30}=0.5626$

c) Let X denote the time between arrivals. Then, X is an exponential random variable with $\lambda=1$ arrivals per hour.

P(X > x) = 0.1 and $P(X>x)=\int_x^\infty 1e^{-1t}dt=-e^{-1t}\Big|_x^\infty=e^{-1x}=0.1$. Therefore, x = 2.3 hours.

4-127. Let X denote the number of calls in 30 minutes. Because the time between calls is an exponential random variable, X is a Poisson random variable with $\lambda=1/E(X)=0.1$ calls per minute. Therefore, E(X) = 3 calls per 30 minutes.

a) P(X > 3) = $1-P(X\le 3)=1-\left[\frac{e^{-3}3^0}{0!}+\frac{e^{-3}3^1}{1!}+\frac{e^{-3}3^2}{2!}+\frac{e^{-3}3^3}{3!}\right]=0.3528$

b) P(X = 0) = $\frac{e^{-3}3^0}{0!}=0.04979$

c) Let Y denote the time between calls in minutes. Then, $P(Y\ge x)=0.01$ and

$P(Y\ge x)=\int_x^\infty 0.1e^{-0.1t}dt=-e^{-0.1t}\Big|_x^\infty=e^{-0.1x}$. Therefore, $e^{-0.1x}=0.01$ and x = 46.05 minutes.

. d) P(Y > 120) = $\int_{120}^\infty 0.1e^{-0.1y}dy=-e^{-0.1y}\Big|_{120}^\infty=e^{-12}=6.14\times10^{-6}$.

e) Because the calls are a Poisson process, the numbers of calls in disjoint intervals are independent. The probability of no calls in one-half hour is $e^{-3}=0.04979$. Therefore, the answer is $\left[e^{-3}\right]^4=e^{-12}=6.14\times10^{-6}$.

Alternatively, the answer is the probability of no calls in two hours. From part d) of this exercise, this is e^{-12}.

f) Because a Poisson process is memoryless, probabilities do not depend on whether or not intervals are consecutive. Therefore, parts d) and e) have the same answer.

4-128. X is an exponential random variable with $\lambda=0.2$ flaws per meter.

a) E(X) = $1/\lambda=5$ meters.

b) P(X > 10) = $\int_{10}^\infty 0.2e^{-0.2x}dx=-e^{-0.2x}\Big|_{10}^\infty=e^{-2}=0.1353$

c) No

d) P(X < x) = 0.90. Then, P(X < x) = $-e^{-0.2t}\Big|_0^x=1-e^{-0.2x}$.

Therefore, $1-e^{-0.2x}=0.9$ and x = 11.51.

P(X > 8) = $\int_8^\infty 0.2e^{-0.2x}dx = -e^{-8/5} = 0.2019$

The distance between successive flaws is either less than 8 meters or not. The distances are independent and P(X > 8) = 0.2019.
Let Y denote the number of flaws until the distance exceeds 8 meters. Then, Y is a geometric random variable with p = 0.2019.

e) P(Y = 5) = $(1-0.2019)^4 0.2019 = 0.0819$

f) E(Y) = 1/0.2019 = 4.95.

4-129. a) $P(X > \theta) = \int_\theta^\infty \frac{1}{\theta}e^{-x/\theta}dx = -e^{-x/\theta}\Big|_\theta^\infty = e^{-1} = 0.3679$

b) $P(X > 2\theta) = -e^{-x/\theta}\Big|_{2\theta}^\infty = e^{-2} = 0.1353$

c) $P(X > 3\theta) = -e^{-x/\theta}\Big|_{3\theta}^\infty = e^{-3} = 0.0498$

d) The results do not depend on θ.

4-130. $E(X) = \int_0^\infty x\lambda e^{-\lambda x}dx.$ Use integration by parts with u = x and dv = $\lambda e^{-\lambda x}$.

Then, $E(X) = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}dx = \frac{-e^{-\lambda x}}{\lambda}\Big|_0^\infty = 1/\lambda$

V(X) = $\int_0^\infty (x-\frac{1}{\lambda})^2 \lambda e^{-\lambda x}dx.$ Use integration by parts with $u = (x-\frac{1}{\lambda})^2$ and

dv = $\lambda e^{-\lambda x}$. Then,

$V(X) = -(x-\frac{1}{\lambda})^2 e^{-\lambda x}\Big|_0^\infty + 2\int_0^\infty (x-\frac{1}{\lambda})e^{-\lambda x}dx = (\frac{1}{\lambda})^2 + \frac{2}{\lambda}\int_0^\infty (x-\frac{1}{\lambda})\lambda e^{-\lambda x}dx$

The last integral is seen to be zero from the definition of E(X). Therefore, V(X) = $\left(\frac{1}{\lambda}\right)^2$.

4-131. X is an exponential random variable with μ = 3.5 days.

a) P(X < 2) = $\int_0^2 \frac{1}{3.5}e^{-x/3.5}dx = 1 - e^{-2/3.5} = 0.435$

b) $P(X > 7) = \int_7^\infty \frac{1}{3.5}e^{-x/3.5}dx = e^{-7/3.5} = 0.135$

c) $P(X > x) = 0.9$ and $P(X > x) = e^{-x/3.5} = 0.9$
Therefore, x = –3.5ln(0.9) = 0.369

d) From the lack of memory property P(X < 10 | X > 3) = P(X < 7) and from part (b) this equals

1 – 0.135 = 0.865

4-132. a) $\mu = E(X) = \dfrac{1}{\lambda} = 4.6$, then $\lambda = 0.2174$

$\sigma = \dfrac{1}{\lambda} = 4.6$

b) $P(X > 10) = \int_{10}^{\infty} \dfrac{1}{4.6} e^{-x/4.6} dx = e^{-10/4.6} \Big| = 0.1137$

c) $P(X > x) = \int_{x}^{\infty} \dfrac{1}{4.6} e^{-u/4.6} du = e^{-x/4.6} = 0.25$

Then, x = -4.6ln(0.25) = 6.38

4-133. a) $P(X > 2) = \int_{2}^{\infty} \lambda \exp(-\lambda x) = \exp(-\dfrac{2}{1.48}) = 0.259$

b) $P(X > 1) = \exp(-\dfrac{1}{1.48}) = 0.509$

c) $P(X > 0.5) = \exp(-\dfrac{0.5}{1.48}) = 0.713$

4-134.

a) Let X_1 denote the time until the first request. Then X_1 is exponentially distributed with $\lambda = 5.0$

$P(X_1 < 0.4) = 1 - \exp(-5 \times 0.4) = 1 - \exp(-2) \cong 0.865$

b) Let X_2 denote the time between the second and the third requests. Then X_2 is exponentially distributed with $\lambda = 5.0$.

$P(X_2 > 7.5) = \exp(-5 \times 7.5) = \exp(-37.5) \cong 0$

c) $P(X_1 > 0.5) = \exp(-0.5\lambda) = 0.9$

$-0.5\lambda = \ln(0.9)$, then $\lambda = 0.21$

d) In the long term, the queue of service requests is expected to continue to increase because the service rate is less than the arrival rate of requests.

4-135. a) $P(X < 5.5) = 1 - \exp(-\dfrac{5.5}{7}) = 0.544$

b) $P(X > 10 \mid X > 7) = P(X > 3) = \exp(-3/7) = 0.651$

$P(X < 6) = 0.9 \Rightarrow 1 - \exp(-\lambda 6) = 0.9 \Rightarrow \lambda = 0.384$

c) $\mu = \dfrac{1}{0.38} = 2.61$

4-136. Let X denote the time between two consecutive arrivals.

a) $\mu = E(X) = \dfrac{1}{\lambda} = \dfrac{1}{2.5} = 0.4$ time units

b) $P(X > 0.3) = \exp(-2.5 \times 0.3) = \exp(-0.75) \cong 0.472$

c) $P(X > 1.0) = \exp(-2.5 \times 1.0) = \exp(-2.5) \cong 0.08$

d) $P(X > 0.3) = \exp(-0.3\lambda) = 0.9$

$-0.3\lambda = \ln(0.9)$, then $\lambda = 0.35$

Section 4-9

4-137. a) $\Gamma(6) = 5! = 120$

b) $\Gamma(\frac{5}{2}) = \frac{3}{2}\Gamma(\frac{3}{2}) = \frac{3}{2}\frac{1}{2}\Gamma(\frac{1}{2}) = \frac{3}{4}\pi^{1/2} = 1.32934$

c) $\Gamma(\frac{9}{2}) = \frac{7}{2}\Gamma(\frac{7}{2}) = \frac{7}{2}\frac{5}{2}\frac{3}{2}\frac{1}{2}\Gamma(\frac{1}{2}) = \frac{105}{16}\pi^{1/2} = 11.6317$

4-138. X is a gamma random variable with the parameters $\lambda = 0.01$ and $r = 3$.
The mean is $E(X) = r/\lambda = 300$.
The variance is $Var(X) = r/\lambda^2 = 30000$.

4-139. a) The time until the tenth call is an Erlang random variable with $\lambda = 5$ calls per minute and r = 10.

b) E(X) = 10/5 = 2 minutes. V(X) = 10/25 = 0.4 minutes.

c) Because a Poisson process is memoryless, the mean time is 1/5 = 0.2 minutes or 12 seconds

Let Y denote the number of calls in one minute. Then, Y is a Poisson random variable with 5 calls per minute.

d) P(Y = 4) = $\dfrac{e^{-5}5^4}{4!} = 0.1755$

e) P(Y > 2) = 1 - $P(Y \le 2) = 1 - \dfrac{e^{-5}5^0}{0!} - \dfrac{e^{-5}5^1}{1!} - \dfrac{e^{-5}5^2}{2!} = 0.8754$

Let W denote the number of one minute intervals out of 10 that contain more than 2 calls. Because the calls are a Poisson process, W is a binomial random variable with n = 10 and p = 0.8754.
Therefore, P(W = 10) = $\binom{10}{10}0.8754^{10}(1-0.8754)^0 = 0.2643$

4-140. Let X denote the pounds of material to obtain 15 particles. Then, X has an Erlang distribution with r = 15 and $\lambda = 0.01$.

a) E(X) = $\dfrac{r}{\lambda} = \dfrac{15}{0.01} = 1500$ pounds.

b) V(X) = $\dfrac{15}{0.01^2} = 150{,}000$ and $\sigma_X = \sqrt{150{,}000} = 387.3$ pounds.

4-141. Let X denote the time between failures of a laser. Here X is exponential with a mean of 25,000.
a) Expected time until the second failure $E(X) = r/\lambda = 2/0.00004 = 50{,}000$ hours
b) N=no of failures in 50000 hours

$$E(N) = \frac{50000}{25000} = 2$$

$$P(N \le 2) = \sum_{k=0}^{2} \frac{e^{-2}(2)^k}{k!} = 0.6767$$

4-142. Let X denote the time until 5 messages arrive at a node. Then, X has an Erlang distribution with r = 5 and $\lambda = 30$ messages per minute.
a) E(X) = 5/30 = 1/6 minute = 10 seconds.
b)V(X) = $\dfrac{5}{30^2} = 1/180$ minute ^{2} = 1/3 second and $\sigma_X = 0.0745$ minute = 4.472 seconds.
c) Let Y denote the number of messages that arrive in 10 seconds. Then, Y is a Poisson random variable with $\lambda = 30$ messages per minute = 5 messages per 10 seconds.

$$P(Y \ge 5) = 1 - P(Y \le 4) = 1 - \left[\tfrac{e^{-5}5^0}{0!} + \tfrac{e^{-5}5^1}{1!} + \tfrac{e^{-5}5^2}{2!} + \tfrac{e^{-5}5^3}{3!} + \tfrac{e^{-5}5^4}{4!}\right]$$

$$= 0.5595$$

d) Let Y denote the number of messages that arrive in 5 seconds. Then, Y is a Poisson random variable with E(Y) =2.5 messages per 5 seconds.

$$P(Y \ge 5) = 1 - P(Y \le 4) = 1 - 0.8912 = 0.1088$$

4-143. Let X denote the number of bits until five errors occur. Then, X has an Erlang distribution with r = 5 and $\lambda = 10^{-5}$ error per bit.
a) E(X) = $\dfrac{r}{\lambda} = 5 \times 10^5$ bits.
b) V(X) = $\dfrac{r}{\lambda^2} = 5 \times 10^{10}$ and $\sigma_X = \sqrt{5 \times 10^{10}} = 223607$ bits.
c) Let Y denote the number of errors in 10^5 bits. Then, Y is a Poisson random variable with $\lambda = 1/10^5 = 10^{-5}$ error per bit = 1 error per 10^5 bits.

$$P(Y \ge 3) = 1 - P(Y \le 2) = 1 - \left[\tfrac{e^{-1}1^0}{0!} + \tfrac{e^{-1}1^1}{1!} + \tfrac{e^{-1}1^2}{2!}\right] = 0.0803$$

4-144. $\lambda = 20 \quad r = 100$
a) $E(X) = r/\lambda = 100/20 = 5$ minutes
b) 4 min - 2.5 min=1.5 min
c) Let Y be the number of calls before 15 seconds. Here E(Y) = 0.25(20) = 5.

$$P(Y \ge 3) = 1 - P(X \le 2) = 1 - \left[\tfrac{e^{-5}5^0}{0!} + \tfrac{e^{-5}5^1}{1!} + \tfrac{e^{-5}5^2}{2!}\right] = 1 - .1247 = 0.8753$$

4-145. a) Let X denote the number of customers that arrive in 10 minutes. Then, X is a Poisson random variable with $\lambda = 0.2$ arrivals per minute = 2 arrivals per 10 minutes.

$$P(X > 3) = 1 - P(X \le 3) = 1 - \left[\tfrac{e^{-2}2^0}{0!} + \tfrac{e^{-2}2^1}{1!} + \tfrac{e^{-2}2^2}{2!} + \tfrac{e^{-2}2^3}{3!}\right] = 0.1429$$

b) Let Y denote the number of customers that arrive in 15 minutes. Then, Y is a Poisson random variable with E(Y) = 3 arrivals per 15 minutes.

$$P(Y \geq 5) = 1 - P(Y \leq 4) = 1 - \left[\frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} + \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!} + \frac{e^{-3}3^4}{4!}\right] = 0.1847$$

4-146. $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx$. Use integration by parts with $u = x^{r-1}$ and $dv = e^{-x}$. Then,

$$\Gamma(r) = -x^{r-1}e^{-x}\Big|_0^\infty + (r-1)\int_0^\infty x^{r-2}e^{-x}dx = (r-1)\Gamma(r-1).$$

4-147. $\int_0^\infty f(x;\lambda,r)dx = \int_0^\infty \frac{\lambda^r x^{r-1}e^{-\lambda x}}{\Gamma(r)}dx$. Let y = λx, then the integral is $\int_0^\infty \frac{\lambda y^{r-1}e^{-y}}{\Gamma(r)}\frac{dy}{\lambda}$. From the definition of $\Gamma(r)$, this integral is recognized to equal 1.

4-148. If X is a chi-square random variable, then X is a special case of a gamma random variable. Now, E(X) = $\frac{r}{\lambda} = \frac{(7/2)}{(1/2)} = 7$ and $V(X) = \frac{r}{\lambda^2} = \frac{(7/2)}{(1/2)^2} = 14$

4-149. Let X denote the number of patients arrive at the emergency department. Then, X has a Poisson distribution with $\lambda = 6.5$ patients per hour.

a) $E(X) = r/\lambda = 10/6.5 =$ 1.539 hour.

b) Let Y denote the number of patients that arrive in 20 minutes. Then, Y is a Poisson random variable with E(Y) = 6.5/3 = 2.1667 arrivals per 20 minutes. The event that the third arrival exceeds 20 minutes is equivalent to the event that there are two or fewer arrivals in 20 minutes. Therefore,

$$P(Y \leq 2) = \left[\frac{e^{-2.1667}2.1667^0}{0!} + \frac{e^{-2.1667}2.1667^1}{1!} + \frac{e^{-2.1667}2.1667^2}{2!}\right] = 0.6317$$

The solution may also be obtained from the result that the time until the third arrival follows a gamma distribution with r = 3 and λ = 6.5 arrivals per hour. The probability is obtained by integrating the probability density function from 20 minutes to infinity.

4-150. a) $E(X) = r/\lambda = 18$, then $r = 18\lambda$

$Var(X) = r/\lambda^2 = 18/\lambda = 36$, then $\lambda = 0.5$

Therefore, the parameters are $\lambda = 0.5$ and $r = 18\lambda = 18(0.5) = 9$

b) The distribution of each step is exponential with λ = 0.5 and 9 steps produce this gamma distribution.

4-151. a) Mean of 1.74 particles per 200 nanoseconds.
Therefore, the mean of 1.74/200 = 0.0087 particles per nanosecond. Here r = 100.

$$\mu = \frac{100}{\lambda} = 11494.253$$

$$\sigma^2 = \frac{100}{\lambda^2} = 1321178.491$$

b) X denotes the time until fifth particle arrives
N denotes the number of particles in 1 nanosecond
N has a Poisson distribution with $\lambda = 0.0087$ particles per nanosecond

$$P(X > 1 \text{ nanosecond}) = P(N \le 4) = \sum_{n=0}^{4} e^{-\lambda} \frac{\lambda^n}{n!}$$
$$= 0.991338 + 0.008625 + 3.75\text{E-}05 + 1.09\text{E-}07 + 2.37\text{E-}10 = 1$$

4-152. Let X denote the time delay.

a) $\mu = E(X) = \dfrac{r}{\lambda} = 4(0.25) = 1$

$$\sigma^2 = V(X) = \frac{r}{\lambda^2} = 4(0.25)^2 = 0.25$$

b) Because $r = 4$ is an integer, we may use the Erlang distribution.

$$P(X > 0.5) = \sum_{k=0}^{r-1} \frac{e^{-0.5\lambda}(0.5\lambda)^k}{k!} = \sum_{k=0}^{3} \frac{e^{-2}2^k}{k!} = 0.857$$

c) $P(0.5 < X < 1.0) = P(X < 1.0) - P(X \le 0.5) = (1 - P(X \ge 1.0)) - (1 - P(X > 0.5))$

$$= \left(1 - \sum_{k=0}^{3} \frac{e^{-4}4^k}{k!}\right) - (1 - 0.857) = (1 - 0.433) - 0.143 = 0.424$$

Section 4-10

4-153.

β=0.2 and δ=100 hours

$$E(X) = 100\Gamma(1 + \tfrac{1}{0.2}) = 100 \times 5! = 12{,}000$$

$$V(X) = 100^2\Gamma(1 + \tfrac{2}{0.2}) - 100^2[\Gamma(1 + \tfrac{1}{0.2})]^2 = 3.61 \times 10^{10}$$

4-154. a) $P(X < 10000) = F_X(10000) = 1 - e^{-100^{0.2}} = 1 - e^{-2.512} = 0.9189$

b) $P(X > 5000) = 1 - F_X(5000) = e^{-50^{0.2}} = 0.1123$

4-155. If X is a Weibull random variable with β =1and δ=1000, the distribution of X is the exponential distribution with λ =0.001.

$$f(x) = \left(\frac{1}{1000}\right)\left(\frac{x}{1000}\right)^0 e^{-\left(\frac{x}{1000}\right)^1} \text{ for } x > 0$$
$$= 0.001e^{-0.001x} \text{ for } x > 0$$

The mean of X is E(X) = 1/λ = 1000.

4-156. Let X denote lifetime of a bearing. β=2 and δ=10000 hours

a) $P(X > 8000) = 1 - F_X(8000) = e^{-\left(\frac{8000}{10000}\right)^2} = e^{-0.8^2} = 0.5273$

b)

$$E(X) = 10000\Gamma(1 + \tfrac{1}{2}) = 10000\Gamma(1.5)$$
$$= 10000(0.5)\Gamma(0.5) = 5000\sqrt{\pi} = 8862.3$$

c) Let Y denote the number of bearings out of 10 that last at least 8000 hours. Then, Y is a binomial random variable with n = 10 and p = 0.5273.

$P(Y = 10) = \binom{10}{10} 0.5273^{10}(1 - 0.5273)^0 = 0.00166$.

4-157. a) $E(X) = \delta\Gamma(1 + \frac{1}{\beta}) = 900\Gamma(1 + 1/3) = 900\Gamma(4/3) = 900(0.89298). = 803.68$ hours

b)

$$V(X) = \delta^2\Gamma(1 + \tfrac{2}{\beta}) - \delta^2\left[\Gamma(1 + \tfrac{2}{\beta})\right]^2 = 900^2\Gamma(1 + \tfrac{2}{3}) - 900^2\left[\Gamma(1 + \tfrac{1}{3})\right]^2$$
$$= 900^2(0.90274)\text{-}900^2(0.89298)^2 = 85319.64 \text{ hours}^2$$

c) $P(X < 500) = F_X(500) = 1 - e^{-\left(\frac{500}{900}\right)^3} = 0.1576$

4-158. Let *X* denote the lifetime.

a) $E(X) = \delta\Gamma(1 + \frac{1}{0.5}) = \delta\Gamma(3) = 2\delta = 600.$ Then $\delta = 300$. Now,

P(X > 500) = $e^{-(\frac{500}{300})^{0.5}} = 0.2750$

b) P(X < 400) = $1 - e^{-(\frac{400}{300})^{0.5}} = 0.6848$

4-159. a) β = 2, δ = 500

$$E(X) = 500\Gamma(1 + \tfrac{1}{2}) = 500\Gamma(1.5)$$
$$= 500(0.5)\Gamma(0.5) = 250\sqrt{\pi} = 443.11$$

b) $V(X) = 500^2\Gamma(1 + 1) - 500^2[\Gamma(1 + \frac{1}{2})]^2$
$$= 500^2\Gamma(2) - 500^2[\Gamma(1.5)]^2 = 53650.5$$

c) P(X < 250) = F(250) = $1 - e^{-(\frac{250}{500})^2} = 1 - 0.7788 = 0.2212$

4-160. $E(X) = \delta\Gamma(1+\frac{1}{2}) = 2.5$

So $\delta = \dfrac{2.5}{\Gamma(1+\frac{1}{2})} = \dfrac{5}{\sqrt{\pi}}$

$Var(X) = \delta^2\Gamma(2) - (EX)^2 = \dfrac{25}{\pi} - 2.5^2 = 1.7077$

Stdev(X)= 1.3068

4-161. $\delta^2\Gamma(1+\frac{2}{\beta}) = Var(X) + (EX)^2 = 10.3 + 4.9^2 = 34.31$

$\delta\Gamma(1+\frac{1}{\beta}) = E(X) = 10.3$

Requires a numerical solution to these two equations.

4-162. a) $P(X < 10) = F_X(10) = 1 - e^{-(10/8.6)^2} = 1 - e^{-1.3521} = 0.7413$

b) $P(X > 9) = 1 - F_X(9) = e^{-(9/8.6)^2} = 0.3345$

c)

$P(8 < X < 11) = F_X(11) - F_X(8) = (1 - e^{-(11/8.6)^2}) - (1 - e^{-(8/8.6)^2}) = 0.8052 - 0.5791 = 0.2261$

d) $P(X > x) = 1 - F_X(x) = e^{-(x/8.6)^2} = 0.9$

Therefore, $-(x/8.6)^2 = \ln(0.9) = -0.1054$, and x = 2.7920

4-163. a) $P(X > 3000) = 1 - F_X(3000) = e^{-(3000/4000)^2} = 0.5698$

b) $P(X > 6000 \mid X > 3000) = \dfrac{P(X > 6000, X > 3000)}{P(X > 3000)} = \dfrac{P(X > 6000)}{P(X > 3000)}$

$= \dfrac{1 - F_X(6000)}{1 - F_X(3000)} = \dfrac{e^{-(6000/4000)^2}}{e^{-(3000/4000)^2}} = \dfrac{0.1054}{0.5698} = 0.1850$

c) If it is an exponential distribution, then β = 1 and

$= \dfrac{1 - F_X(6000)}{1 - F_X(3000)} = \dfrac{e^{-(6000/4000)}}{e^{-(3000/4000)}} = \dfrac{0.2231}{0.4724} = 0.4724$

For the Weibull distribution (with β = 2) there is no lack of memory property so that the answers to parts (a) and (b) differ whereas they would be the same if an exponential distribution were assumed. From part (b), the probability of survival beyond 6000 hours, given the device has already survived 3000 hours, is lower than the probability of survival beyond 3000 hours from the start time.

4-164. a) $P(X > 3500) = 1 - F_X(3500) = e^{-(3500/4000)^{0.5}} = 0.4206$

b) $P(X > 6000 \mid X > 3000) = \dfrac{P(X > 6000, X > 3000)}{P(X > 3000)} = \dfrac{P(X > 6000)}{P(X > 3000)}$

$$= \frac{1 - F_X(6000)}{1 - F_X(3000)} = \frac{e^{-(6000/4000)^{0.5}}}{e^{-(3000/4000)^{0.5}}} = \frac{0.2938}{0.4206} = 0.6986$$

c) $P(X > 6000 \mid X > 3000) = \dfrac{P(X > 6000, X > 3000)}{P(X > 3000)} = \dfrac{P(X > 6000)}{P(X > 3000)}$

If it is an exponential distribution, then β = 1

$$= \frac{1 - F_X(6000)}{1 - F_X(3000)} = \frac{e^{-(6000/4000)}}{e^{-(3000/4000)}} = \frac{0.2231}{0.4724} = 0.4724$$

For the Weibull distribution (with β = 0.5) there is no lack of memory property so that the answers to parts (a) and (b) differ whereas they would be the same if an exponential distribution were assumed. From part (b), the probability of survival beyond 6000 hours, given the device has already survived 3000 hours, is greater than the probability of survival beyond 3000 hours from the start time.

d) The failure rate can be increased or decreased relative to the exponential distribution with the shape parameter β in the Weibull distribution.

4-165. a) $P(X > 3500) = 1 - F_X(3500) = e^{-(3500/2000)^2} = 0.0468$

b) The mean of this Weibull distribution is (2000) $0.5\sqrt{\pi}$ = 1772.45

If it is an exponential distribution with this mean then

$$P(X > 3500) = 1 - F_X(3500) = e^{-(3500/1772.45)} = 0.1388$$

c) The probability that the lifetime exceeds 3500 hours is greater under the exponential distribution than under this Weibull distribution model.

4-166. a) $\mu = E(X) = \delta\Gamma\left(1 + \frac{1}{\beta}\right) = 16.01$

$$\sigma^2 = V(X) = \delta^2\Gamma\left(1 + \frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2 = 11.66^2$$

$$\delta^2\Gamma\left(1 + \frac{2}{\beta}\right) = 11.66^2 + \delta^2\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2 = 11.66^2 + 16.01^2 = 392.2757$$

$$\delta^2\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2 = 16.01^2 = 256.3201$$

$$\frac{\delta^2\Gamma\left(1+\frac{2}{\beta}\right)}{\delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2}=\frac{\frac{2}{\beta}\Gamma\left(\frac{2}{\beta}\right)}{\left(\frac{1}{\beta}\right)^2\left[\Gamma\left(\frac{1}{\beta}\right)\right]^2}=\frac{2\Gamma\left(\frac{2}{\beta}\right)}{\left(\frac{1}{\beta}\right)\left[\Gamma\left(\frac{1}{\beta}\right)\right]^2}=\frac{392.2757}{256.3201}\text{, then } \beta\cong 1.39\,.$$

$\mu=\delta\Gamma\left(1+\frac{1}{1.39}\right)=16.01$, then $\delta\cong 17.55$.

We have used the property of the gamma function $\Gamma(r)=(r-1)\Gamma(r-1)$ in this solution.

b) Let X denote the survival time.

$$P(X>48)=1-P(X\le 48)=1-F(48)=1-\left[1-\exp\left(-\left(\frac{48}{17.55}\right)^{1.39}\right)\right]=0.017$$

c) Find a such that $P(X>a)=0.90$.

$$P(X>a)=1-P(X\le a)=1-F(a)=1-\left[1-\exp\left(-\left(\frac{a}{17.55}\right)^{1.39}\right)\right]=0.90$$

Then, $a=3.477$

4-167.

a) $\mu=300$
$\sigma^2=90000\Rightarrow\sigma=300$

b) $P(X>240)=\exp(-\left(\frac{240}{\delta}\right)^{\beta})=0.449$

c) $P(X>a)=\exp(-\left(\frac{a}{300}\right)^{1})=0.25\Rightarrow a=415.88$

4-168. Let X denote the average annual losses (in billions of dollars)

a) $P(X>2)=1-P(X\le 2)=1-F(2)=1-\left[1-\exp\left(-\left(\frac{2}{1.9317}\right)^{0.8472}\right)\right]=0.357$

b)

$$P(2<X<4)=P(X<4)-P(X<2)=\left[1-\exp\left(-\left(\frac{4}{1.9317}\right)^{0.8472}\right)\right]-(1-0.357)=0.200$$

c) Find a such that $P(X>a)=0.05$

$$P(X>a)=1-P(X\le a)=1-F(a)=\exp\left(-\left(\frac{a}{1.9317}\right)^{0.8472}\right)=0.05$$

Then, $a=7.053$

d) $\mu = E(X) = \delta\Gamma\left(1+\frac{1}{\beta}\right) = (1.9317)\Gamma\left(1+\frac{1}{0.8472}\right) = 2.106$

$$\sigma = \sqrt{V(X)} = \sqrt{\delta^2\Gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2}$$

$$= \sqrt{(1.9317)^2\Gamma\left(1+\frac{2}{0.8472}\right) - (1.9317)^2\left[\Gamma\left(1+\frac{1}{0.8472}\right)\right]^2} = 2.497$$

4-169.

$$\begin{aligned} &P(X < 26) = 0.01 \\ &P(X < 31.6) = 0.07 \end{aligned} \Rightarrow$$

$$\left.\begin{aligned} &1-\exp(-(\frac{26}{\delta})^{\beta}) = 0.01 \Rightarrow \left(\frac{26}{\delta}\right)^{\beta} = -\ln(0.99) \Rightarrow \beta\ln\left(\frac{26}{\delta}\right) = \ln(-\ln(0.99)) \\ &1-\exp(-(\frac{31.6}{\delta})^{\beta}) = 0.07 \Rightarrow \left(\frac{31.6}{\delta}\right)^{\beta} = -\ln(0.93) \Rightarrow \beta\ln\left(\frac{31.6}{\delta}\right) = \ln(-\ln(0.93)) \end{aligned}\right. \Rightarrow$$

$$\delta = 40.93$$
$$\beta = 10.13$$

<u>Section 4-11</u>

4-170. X is a lognormal distribution with θ=5 and ω^2=9

a) $P(X < 13300) = P(e^W < 13300) = P(W < \ln(13300)) = \Phi\left(\frac{\ln(13330)-5}{3}\right)$

$$= \Phi(1.50) = 0.9332$$

b) Find the value for which P(X≤x)=0.95

$$P(X \le x) = P(e^W \le x) = P(W < \ln(x)) = \Phi\left(\frac{\ln(x)-5}{3}\right) = 0.95$$

$$\frac{\ln(x)-5}{3} = 1.65 \quad x = e^{1.65(3)+5} = 20952.2$$

c) $\mu = E(X) = e^{\theta+\omega^2/2} = e^{5+9/2} = e^{9.5} = 13359.7$

$$V(X) = e^{2\theta+\omega^2}(e^{\omega^2}-1) = e^{10+9}(e^9-1) = e^{19}(e^9-1) = 1.45x10^{12}$$

4-171. a) X is a lognormal distribution with θ=-2 and ω^2=9

$$P(500 < X < 1000) = P(500 < e^W < 1000) = P(\ln(500) < W < \ln(1000))$$

$$= \Phi\left(\frac{\ln(1000)+2}{3}\right) - \Phi\left(\frac{\ln(500)+2}{3}\right) = \Phi(2.97) - \Phi(2.74) = 0.0016$$

b) $P(X < x) = P(e^W \le x) = P(W < \ln(x)) = \Phi\left(\frac{\ln(x)+2}{3}\right) = 0.1$

$$\frac{\ln(x)+2}{3} = -1.28 \qquad x = e^{-1.28(3)-2} = 0.0029$$

c) $\mu = E(X) = e^{\theta+\omega^2/2} = e^{-2+9/2} = e^{2.5} = 12.1825$

$$V(X) = e^{2\theta+\omega^2}(e^{\omega^2}-1) = e^{-4+9}(e^9-1) = e^5(e^9-1) = 1{,}202{,}455.87$$

4-172. a) X is a lognormal distribution with θ=2 and $\omega^2=4$

$$P(X < 500) = P(e^W < 500) = P(W < \ln(500)) = \Phi\left(\frac{\ln(500)-2}{2}\right)$$
$$= \Phi(2.11) = 0.9826$$

b)

$$P(X < 15000 \mid X > 1000) = \frac{P(1000 < X < 1500)}{P(X > 1000)}$$
$$= \frac{\left[\Phi\left(\frac{\ln(1500)-2}{2}\right) - \Phi\left(\frac{\ln(1000)-2}{2}\right)\right]}{\left[1-\Phi\left(\frac{\ln(1000)-2}{2}\right)\right]}$$
$$= \frac{\Phi(2.66)-\Phi(2.45)}{\left(1-\Phi(2.45)\right)} = \frac{0.9961-0.9929}{\left(1-0.9929\right)} = 0.0032/0.007 = 0.45$$

c) The product has degraded over the first 1000 hours, so the probability of it lasting another 500 hours is very low.

4-173. X is a lognormal distribution with θ=0.5 and $\omega^2=1$

a)

$$P(X > 10) = P(e^W > 10) = P(W > \ln(10)) = 1-\Phi\left(\frac{\ln(10)-0.5}{1}\right)$$
$$= 1-\Phi(1.80) = 1-0.96407 = 0.03593$$

b) $P(X \le x) = P(e^W \le x) = P(W < \ln(x)) = \Phi\left(\frac{\ln(x)-0.5}{1}\right) = 0.50$

$\frac{\ln(x)-0.5}{1} = 0 \quad x = e^{0(1)+0.5} = 1.65$ seconds

c) $\mu = E(X) = e^{\theta+\omega^2/2} = e^{0.5+1/2} = e^1 = 2.7183$

$$V(X) = e^{2\theta+\omega^2}(e^{\omega^2}-1) = e^{1+1}(e^1-1) = e^2(e^1-1) = 12.6965$$

4-174. Find the values of θ and ω^2 given that E(X) = 100 and V(X) = 85,000

$$100 = e^{\theta+\omega^2/2} \qquad 85000 = e^{2\theta+\omega^2}(e^{\omega^2} - 1)$$

Let $x = e^{\theta}$ and $y = e^{\omega^2}$ then

$100 = x\sqrt{y}$ and $85000 = x^2 y(y-1) = x^2y^2 - x^2 y$

Square the first equation to obtain $10000 = x^2 y$ and substitute into the second equation

$$85000 = 10000(y-1)$$

$$y = 9.5$$

Substitute y into the first equation and solve for x to obtain

$$x = \frac{100}{\sqrt{9.5}} = 32.444$$

$\theta = \ln(32.444) = 3.48$ and $\omega^2 = \ln(9.5) = 2.25$

4-175. a) Find the values of θ and ω^2 given that E(X) = 10000 and σ = 20,000

$$10000 = e^{\theta+\omega^2/2} \qquad 20000^2 = e^{2\theta+\omega^2}(e^{\omega^2} - 1)$$

Let $x = e^{\theta}$ and $y = e^{\omega^2}$ then

$10000 = x\sqrt{y}$ and $20000^2 = x^2 y(y-1) = x^2y^2 - x^2 y$

Square the first equation $10000^2 = x^2 y$ and substitute into the second equation

$$20000^2 = 10000^2(y-1)$$

$$y = 5$$

Substitute y into the first equation and solve for x to obtain $x = \dfrac{10000}{\sqrt{5}} = 4472.1360$

$\theta = \ln(4472.1360) = 8.4056$ and $\omega^2 = \ln(5) = 1.6094$

b)

$$P(X > 10000) = P(e^W > 10000) = P(W > \ln(10000)) = 1 - \Phi\left(\frac{\ln(10000) - 8.4056}{1.2686}\right)$$

$$= 1 - \Phi(0.63) = 1 - 0.7357 = 0.2643$$

c) $P(X > x) = P(e^W > x) = P(W > \ln(x)) = \Phi\left(\dfrac{\ln(x) - 8.4056}{1.2686}\right) = 0.1$

$$\frac{\ln(x) - 8.4056}{1.2686} = -1.28 \quad x = e^{-1.280(1.2686)+8.4056} = 881.65 \text{ hours}$$

4-176. $E(X) = \exp(\theta + \omega^2/2) = 120.87$

$$\sqrt{\exp(\omega^2)-1} = 0.09$$

So

$$\omega = \sqrt{\ln 1.0081} = 0.0898 \text{ and}$$

$$\theta = \ln 120.87 - \omega^2/2 = 4.791$$

4-177. Let X ~ N(μ, σ^2), then Y = e^X follows a lognormal distribution with mean μ and variance σ^2. By definition, $F_Y(y) = P(Y \le y) = P(e^X < y) = P(X < \log y) = F_X(\log y) = \Phi\left(\frac{\log y - \mu}{\sigma}\right)$.

Because Y = e^X and X ~ N(μ, σ^2), we can show that $f_Y(Y) = \frac{1}{y} f_X(\log y)$

Finally, $f_Y(y) = \dfrac{\partial F_Y(y)}{\partial y} = \dfrac{\partial F_X(\log y)}{\partial y} = \dfrac{1}{y} f_X(\log y) = \dfrac{1}{y}\cdot\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{\log y - \mu}{2\sigma}\right)^2}$.

4-178. X has a lognormal distribution with θ = 10 and ω^2 = 16

a) $P(X < 2000) = P(e^W < 2000) = P(W < \ln(2000)) = \Phi\left(\dfrac{\ln(2000)-10}{4}\right)$

$$= \Phi(-0.5998) = 0.2743$$

b)

$$P(X > 1500) = 1 - P(e^W < 1500) = 1 - P(W < \ln(1500)) = \Phi\left(\frac{\ln(1500)-10}{4}\right)$$

$$= 1 - \Phi(-0.6717) = 1 - 0.2509 = 0.7491$$

c)

$$P(X > x) = P(e^W > x) = P(W > \ln(x)) = 1 - \Phi\left(\frac{\ln(x)-10}{4}\right) = 0.7$$

$$-0.5244 = \frac{\ln(x)-10}{4}$$

Therefore, $x = 2703.76$

4-179. X has a lognormal distribution with θ = 1.5 and ω = 0.4

a) $\mu = E(X) = e^{\theta+\omega^2/2} = e^{1.5+0.16/2} = e^{1.58} = 4.8550$

$$V(X) = e^{2\theta+\omega^2}(e^{\omega^2}-1) = e^{3+0.16}(e^{0.16}-1) = 4.0898$$

b) $P(X < 8) = P(e^W < 8) = P(W < \ln(8)) = \Phi\left(\dfrac{\ln(8)-1.5}{0.4}\right) = \Phi(1.4486) = 0.9263$

c) $P(X < 0) = 0$ for the lognormal distribution. If the distribution is normal, then

$$P(X < 0) = P(Z < \frac{0-4.855}{\sqrt{4.0898}}) = 0.008$$

Because waiting times cannot be negative, the normal distribution generates some modeling error.

4-180. Let X denote the particle-size distribution (in centimeters).

a) $P(X < 0.02) = P\left[Z \le \dfrac{\ln(0.02) - \theta}{\omega}\right] = P\left[Z \le \dfrac{\ln(0.02) - (-3.8)}{0.7}\right] = \Phi[-0.16] = 0.436$

b) Find x such that $P(X \le x) = 0.95$

$P(X \le x) = P\left[Z \le \dfrac{\ln(x) - (-3.8)}{0.7}\right] = 0.95$

$\dfrac{\ln(x) - (-3.8)}{0.7} = 1.645$, then $x \cong 0.071$

c) $\mu = E(X) = \exp\left(\theta + \omega^2 / 2\right) = 0.0286$

$\sigma^2 = V(X) = \exp\left(2\theta + \omega^2\right)\left(\exp\left(\omega^2\right) - 1\right) = 0.0005$

4-181. a) $P(X > 5) = 1 - P(W < \ln(5)) = 1 - \Phi(\dfrac{\ln(5) - 1}{1}) = 0.271$

b)

$P(X < 8 \mid X > 5) = \dfrac{P(5 < X < 8)}{P(X > 5)} = \dfrac{P(X < 8) - P(X < 5)}{P(X > 5)} =$

$\dfrac{[\Phi(\ln(8) - 1)] - [\Phi(\ln(5) - 1)]}{1 - \Phi(\ln(5) - 1)} = \dfrac{0.86 - 0.73}{1 - 0.73} = 0.483$

c)

$\mu = \exp(1 + 0.5) = 4.48$

$\sigma^2 = \exp(3)(\exp(1) - 1) = 34.517$

4-182. Let X denote the levels of 2,3,7,8-TCDD in human adipose tissue.

$\mu = E(X) = \exp\left(\theta + \omega^2 / 2\right) = 8$

$\sigma^2 = V(X) = \exp\left(2\theta + \omega^2\right)\left(\exp\left(\omega^2\right) - 1\right) = 21$

$\sigma^2 = \left[\exp\left(\theta + \omega^2 / 2\right)\right]^2\left(\exp\left(\omega^2\right) - 1\right) = \mu^2\left(\exp\left(\omega^2\right) - 1\right) = 21$

Then, $\exp\left(\omega^2\right) - 1 = \dfrac{21}{64}$ and $\omega = 0.5327$

$\exp\left(\theta + 0.5327^2 / 2\right) = 8$, then $\theta = 1.9376$

a) $P(2000 < X < 2500) = P(X < 2500) - P(X \le 2000)$

$= P\left[Z < \dfrac{\ln(2500) - \theta}{\omega}\right] - P\left[Z \le \dfrac{\ln(2000) - \theta}{\omega}\right] = \Phi[11.05] - \Phi[10.63] = 0$

b) Find a such that $P(X > a) = 0.10$

$$P(X > a) = 1 - P\left[Z \le \frac{\ln(a) - 1.9376}{0.5327}\right] = 0.10$$

$\dfrac{\ln(a) - 1.9376}{0.5327} = 1.282$, then $a = 13.736$.

c) $\mu = E(X) = \exp\left(\theta + \omega^2/2\right) = 8$

$$\sigma^2 = V(X) = \exp\left(2\theta + \omega^2\right)\left(\exp\left(\omega^2\right) - 1\right) = 21$$

4-183.

a) $P(X < 1000) = \Phi(\dfrac{\ln(1000) - 10}{1.5}) = 0.0196$

$$P(X < 11{,}000 \mid X > 10{,}000) = \frac{P(10000 < X < 11000)}{P(X > 10000)}$$

b)
$$= \frac{\Phi(\frac{\ln(11{,}000) - 10}{1.5}) - \Phi(\frac{\ln(10{,}000) - 10}{1.5})}{1 - \Phi(\frac{\ln(10{,}000) - 10}{1.5})} = 0.032$$

Section 4-12

4-184. The probability density is symmetric.

4-185. a) $P(X < 0.25) = \displaystyle\int_0^{0.25} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}) x^{\alpha - 1}(1 - x)^{\beta - 1}$

$$= \int_0^{0.25} \frac{\Gamma(3.5)}{\Gamma(2.5)\Gamma(1)}) x^{1.5} = \frac{(2.5)(1.5)(0.5)\sqrt{\pi}}{(1.5)(0.5)\sqrt{\pi}} \left.\frac{x^{2.5}}{2.5}\right|_0^{0.25} = 0.25^{2.5} = 0.0313$$

b) $P(0.25 < X < 0.75) = \displaystyle\int_{0.25}^{0.75} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}) x^{\alpha - 1}(1 - x)^{\beta - 1}$

$$= \int_{0.25}^{0.75} \frac{\Gamma(3.5)}{\Gamma(2.5)\Gamma(1)}) x^{1.5} = \frac{(2.5)(1.5)(0.5)\sqrt{\pi}}{(1.5)(0.5)\sqrt{\pi}} \left.\frac{x^{2.5}}{2.5}\right|_{0.25}^{0.75} = 0.75^{2.5} - 0.25^{2.5} = 0.4559$$

c) $\mu = E(X) = \dfrac{\alpha}{\alpha + \beta} = \dfrac{2.5}{2.5 + 1} = 0.7143$

$$\sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{2.5}{(3.5)^2(4.5)} = 0.0454$$

4-186. a) $P(X < 0.25) = \displaystyle\int_0^{0.25} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}) x^{\alpha - 1}(1 - x)^{\beta - 1}$

$$= \int_0^{0.25} \frac{\Gamma(5.2)}{\Gamma(1)\Gamma(4.2)})(1-x)^{3.2} = \frac{(4.2)(3.2)(2.2)(1.2)\Gamma(1.2)}{(3.2)(2.2)(1.2)\Gamma(1.2)} \frac{(-1)(1-x)^{4.2}}{4.2}\Bigg|_0^{0.25} = -(0.75)^{4.2}+1 = 0.7013$$

b) $P(0.5 < X) = \int_{0.5}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)})x^{\alpha-1}(1-x)^{\beta-1}$

$$= \int_{0.5}^{1} \frac{\Gamma(5.2)}{\Gamma(1)\Gamma(4.2)})(1-x)^{3.2} = \frac{(4.2)(3.2)(2.2)(1.2)\Gamma(1.2)}{(3.2)(2.2)(1.2)\Gamma(1.2)} \frac{(-1)(1-x)^{4.2}}{4.2}\Bigg|_{0.5}^{1} = 0+(0.5)^{4.2} = 0.0544$$

c) $\mu = E(X) = \frac{\alpha}{\alpha+\beta} = \frac{1}{1+4.2} = 0.1923$

$$\sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{4.2}{(5.2)^2(6.2)} = 0.0251$$

4-187. a) Mode = $\frac{\alpha-1}{\alpha+\beta-2} = \frac{2}{3+1.4-2} = 0.8333$

$$\mu = E(X) = \frac{\alpha}{\alpha+\beta} = \frac{3}{3+1.4} = 0.6818$$

$$\sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{4.2}{(4.4)^2(5.4)} = 0.0402$$

b) Mode = $\frac{\alpha-1}{\alpha+\beta-2} = \frac{9}{10+6.25-2} = 0.6316$

$$\mu = E(X) = \frac{\alpha}{\alpha+\beta} = \frac{10}{10+6.25} = 0.6154$$

$$\sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{62.5}{(16.25)^2(17.25)} = 0.0137$$

c) Both the mean and variance from part a) are greater than for part b).

4-188. a) $P(X > 0.9) = \int_{0.9}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)})x^{\alpha-1}(1-x)^{\beta-1}$

$$= \int_{0.9}^{1} \frac{\Gamma(11)}{\Gamma(10)\Gamma(1)})x^9 = \frac{(10)(9)\Gamma(9)}{(9)\Gamma(9)} \frac{x^{10}}{10}\Bigg|_{0.9}^{1} = 1-(0.9^{10}) = 0.6513$$

b) $P(X < 0.5) = \int_{0}^{0.5} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)})x^{\alpha-1}(1-x)^{\beta-1}$

$$= \int_{0}^{0.5} \frac{\Gamma(11)}{\Gamma(10)\Gamma(1)})x^9 = \frac{(10)(9)\Gamma(9)}{(9)\Gamma(9)} \frac{x^{10}}{10}\Bigg|_{0}^{0.5} = 0.5^{10} = 0.0010$$

c) $\mu = E(X) = \frac{\alpha}{\alpha+\beta} = \frac{10}{10+1} = 0.9091$

$$\sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{10}{(11)^2(12)} = 0.0069$$

4-189. Let X denote the completion proportion of the maximum time.
The exercise considers the proportion 2/2.5 = 0.8

$$P(X > 0.8) = \int_{0.8}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)})x^{\alpha-1}(1-x)^{\beta-1}$$

$$= \int_{0.8}^{1} \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)})x(1-x)^2 = \frac{(4)(3)\Gamma(3)}{\Gamma(2)\Gamma(3)}(\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4})\Big|_{0.8}^{1} = 12(0.0833 - 0.0811) = 0.0272$$

4-190. $\mu = E(X) = \dfrac{\alpha}{\alpha+\beta} = 0.3$, then $\beta = 2.33\alpha$

$$\sigma^2 = V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.17^2$$

$$\sigma^2 = V(X) = \frac{\alpha(2.33\alpha)}{(\alpha+2.33\alpha)^2(\alpha+2.33\alpha+1)} = 0.17^2$$, then $\alpha = 1.9$ and $\beta = 4.427$

4-191.

a)
$$\alpha = \frac{(\mu-a)(2m-a-b)}{(m-\mu)(b-a)} = \frac{(1.333-1)(2(1.25)-1-2)}{(1.25-1.333)(2-1)} = 2$$

$$\beta = \frac{\alpha(b-\mu)}{\mu-a} = \frac{2(2-1.333)}{1.333-1} = 4$$

b) $$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{2(4)}{6^2 7} = 0.032 \Rightarrow \sigma = \sqrt{0.032} = 0.178$$

Supplemental Exercises

4-192. $f(x) = 0.04$ for 50< x <75

a) $P(X > 70) = \int_{70}^{75} 0.04dx = 0.2x\Big|_{70}^{75} = 0.2$

b) $P(X < 60) = \int_{50}^{60} 0.04dx = 0.04x\Big|_{50}^{60} = 0.4$

c) $E(X) = \dfrac{75+50}{2} = 62.5$ seconds

$V(X) = \dfrac{(75-50)^2}{12} = 52.0833$ seconds^2

4-193. a) P(X < 40) = $P\left(Z < \frac{40-35}{2}\right)$ = P(Z < 2.5) = 0.99379

b) P(X < 30) = $P\left(Z < \frac{30-35}{2}\right)$ = P(Z < –2.5) = 0.00621

Here 0.621% are scrapped

4-194. a) P(X <45) = $P\left(Z < \frac{45-60}{5}\right)$ = P(Z < -3) = 0.00135

b) P(X > 65) = $P\left(Z > \frac{65-60}{5}\right)$ = P(Z >1) = 1- P(Z < 1)

= 1 - 0.841345= 0.158655

c) P(X < x) = $P\left(Z < \frac{x-60}{5}\right)$ = 0.99

Therefore, $\frac{x-60}{5}$ = 2.33 and x = 72

4-195. a) P(X > 90.3) + P(X < 89.7)

= $P\left(Z > \frac{90.3-90.2}{0.1}\right) + P\left(Z < \frac{89.7-90.2}{0.1}\right)$ = P(Z > 1) + P(Z < –5)

= 1 – P(Z < 1) + P(Z < –5) =1 – 0.84134 +0 = 0.15866

b) The process mean should be set at the center of the specifications; that is, at 90.0.

c) P(89.7 < X < 90.3) = $P\left(\frac{89.7-90}{0.1} < Z < \frac{90.3-90}{0.1}\right)$ = P(–3 < Z < 3) = 0.9973.

The yield is 100(0.9973) = 99.73%

d) P(89.7 < X < 90.3) = $P\left(\frac{89.7-90}{0.1} < Z < \frac{90.3-90}{0.1}\right)$ = P(–3 < Z < 3) = 0.9973

P(X=10) = $(0.9973)^{10}$ = 0.9733

e) Let *Y* represent the number of cases out of the sample of 10 that are between 89.7 and 90.3 ml. Then Y follows a binomial distribution with *n*=10 and *p*=0.9973. Thus, E(*Y*)= 9.973.

4-196. a) P(50 < X < 80) = $P\left(\frac{50-100}{20} < Z < \frac{80-100}{20}\right)$ = P(–2.5 < Z < -1)

= P(Z < –1) – P(Z < –2.5) = 0.15245.

b) P(X > x) = 0.10. Therefore, $P\left(Z > \frac{x-100}{20}\right)$ = 0.10 and $\frac{x-100}{20}$ = 1.28

Therefore, x = 125.6 hours

4-197. E(X) = 1000(0.2) = 200 and V(X) = 1000(0.2)(0.8) = 160

a)

$$P(X > 225) = P(X \geq 226) \cong 1 - P(Z \leq \tfrac{225.5-200}{\sqrt{160}}) = 1 - P(Z \leq 2.02) = 1 - 0.9783 = 0.0217$$

b)

$$P(175 \leq X \leq 225) \cong P(\tfrac{174.5-200}{\sqrt{160}} \leq Z \leq \tfrac{225.5-200}{\sqrt{160}}) = P(-2.02 \leq Z \leq 2.02)$$

$$= 0.9783 - 0.0217 = .9566$$

c) If P(X > x) = 0.01, then $P\left(Z > \dfrac{x-200}{\sqrt{160}}\right) = 0.01$.

Therefore, $\frac{x-200}{\sqrt{160}} = 2.33$ and $x = 229.5$

4-198. The time to failure (in hours) for a laser in a cytometry machine is modeled by an exponential distribution with 0.00004.

a) $P(X > 20{,}000) = \int_{20000}^{\infty} 0.00004 e^{-.00004x}\,dx = -e^{-0.00004x}\Big|_{20000}^{\infty} = e^{-0.8} = 0.4493$

b) $P(X < 30{,}000) = \int_{30000}^{\infty} 0.00004 e^{-.00004x}\,dx = -e^{-0.00004x}\Big|_{0}^{30000} = 1 - e^{-1.2} = 0.6988$

c)

$$P(20{,}000 < X < 30{,}000) = \int_{20000}^{30000} 0.00004 e^{-.0.00004x}\,dx$$

$$= -e^{-0.00004x}\Big|_{20000}^{30000} = e^{-0.8} - e^{-1.2} = 0.1481$$

4-199. Let X denote the number of calls in 3 hours. Because the time between calls is an exponential random variable, the number of calls in 3 hours is a Poisson random variable. Now, the mean time between calls is 0.5 hours and $\lambda = 1/0.5 = 2$ calls per hour. Therefore, E(X) = 6 calls in 3 hours.

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - \left[\frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \frac{e^{-6}6^2}{2!} + \frac{e^{-6}6^3}{3!}\right] = 0.8488$$

4-200. Let X denote the time in days until the fourth problem. Then, X has an Erlang distribution with r = 4 and $\lambda = 1/30$ problem per day.

a) E(X) = $\dfrac{4}{30^{-1}} = 120$ days.

b) Let Y denote the number of problems in 120 days. Then, Y is a Poisson random variable with E(Y) = 4 problems per 120 days.

$$P(Y < 4) = \left[\frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!} + \frac{e^{-4}4^3}{3!}\right] = 0.4335$$

4-201. Let X denote the lifetime

a) $E(X) = 700\Gamma(1 + \frac{1}{2}) = 620.4$

b)

$$V(X) = 700^2\Gamma(2) - 700^2[\Gamma(1.5)]^2$$
$$= 700^2(1) - 700^2(0.25\pi) = 105{,}154.9$$

c) P(X > 620.4) = $e^{-\left(\frac{620.4}{700}\right)^2} = 0.4559$

4-202. a) $E(X) = \exp(\theta + \omega^2/2) = 0.001$

$\sqrt{\exp(\omega^2) - 1} = 2$

So

$\omega = \sqrt{\ln 5} = 1.2686$

And

$\theta = \ln 0.001 - \omega^2/2 = -7.7124$

b)

$P(X > 0.005) = 1 - P(\exp(W) \le 0.005) = 1 - P(W \le \ln 0.005)$

$= 1 - \Phi\left(\dfrac{\ln 0.005 + 7.7124}{1.2686}\right) = 0.0285$

4-203. a) $P(X < 2.5) = \int_2^{2.5} (0.5x - 1)dx = \left(0.5\dfrac{x^2}{2} - x\right)\Bigg|_2^{2.5} = 0.0625$

b) $P(X > 3) = \int_3^4 (0.5x - 1)dx = 0.5\frac{x^2}{2} - x\Big|_3^4 = 0.75$

c) $P(2.5 < X < 3.5) = \int_{2.5}^{3.5} (0.5x - 1)dx = 0.5\frac{x^2}{2} - x\Big|_{2.5}^{3.5} = 0.5$

d) $F(x) = \int_2^x (0.5t - 1)dt = 0.5\frac{t^2}{2} - t\Big|_2^x = \frac{x^2}{4} - x + 1$. Then,

$$F(x) = \begin{cases} 0, & x < 2 \\ \frac{x^2}{4} - x + 1, & 2 \le x < 4 \\ 1, & 4 \le x \end{cases}$$

e) $E(X) = \int_2^4 x(0.5x - 1)dx = 0.5\frac{x^3}{3} - \frac{x^2}{2}\Big|_2^4 = \frac{32}{3} - 8 - (\frac{4}{3} - 2) = \frac{10}{3}$

$$V(X) = \int_2^4 (x - \tfrac{10}{3})^2 (0.5x - 1)dx = \int_2^4 (x^2 - \tfrac{20}{3}x + \tfrac{100}{9})(0.5x - 1)dx$$
$$= \int_2^4 (0.5x^3 - \tfrac{13}{3}x^2 + \tfrac{110}{9}x - \tfrac{100}{9})dx = \tfrac{x^4}{8} - \tfrac{13}{9}x^3 + \tfrac{55}{9}x^2 - \tfrac{100}{9}x\Big|_2^4$$
$$= 0.2222$$

4-204. Let X denote the time between calls. Then, $\lambda = 1/E(X) = 0.1$ calls per minute.

a) $P(X < 5) = \int_0^5 0.1e^{-0.1x}dx = -e^{-0.1x}\Big|_0^5 = 1 - e^{-0.5} = 0.3935$

b) $P(5 < X < 15) = -e^{-0.1x}\Big|_5^{15} = e^{-0.5} - e^{-1.5} = 0.3834$

c) P(X < x) = 0.9. Then, $P(X < x) = \int_0^x 0.1e^{-0.1t}dt = 1 - e^{-0.1x} = 0.9$. Now, x = 23.03 minutes.

d) This answer is the same as part a).

$$P(X < 5) = \int_0^5 0.1e^{-0.1x}dx = -e^{-0.1x}\Big|_0^5 = 1 - e^{-0.5} = 0.3935$$

e) This is the probability that there are no calls over a period of 5 minutes. Because a Poisson process is memoryless, it does not matter whether or not the intervals are consecutive.

$$P(X > 5) = \int_5^\infty 0.1e^{-0.1x}dx = -e^{-0.1x}\Big|_5^\infty = e^{-0.5} = 0.6065$$

f) Let Y denote the number of calls in 30 minutes.
Then, Y is a Poisson random variable with E(Y) = 3.

$$P(Y \le 2) = \frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} + \frac{e^{-3}3^2}{2!} = 0.423.$$

g) Let W denote the time until the fifth call. Then, W has an Erlang distribution with $\lambda = 0.1$ and r = 5. E(W) = 5/0.1 = 50 minutes.

4-205. Let X denote the lifetime. Then $\lambda = 1/E(X) = 1/6$.

a) $P(X < 3) = \int_0^3 \tfrac{1}{6}e^{-x/6}dx = -e^{-x/6}\Big|_0^3 = 1 - e^{-0.5} = 0.3935$.

b) Let W denote the number of CPUs that fail within the next three years. Then, W is a binomial random variable with n = 10 and p = 0.3935. Then,

$$P(W \ge 1) = 1 - P(W = 0) = 1 - \binom{10}{0}0.3935^0(1 - 0.3935)^{10} = 0.9933.$$

4-206. X is a lognormal distribution with $\theta=0$ and $\omega^2=4$

a)

$$P(10 < X < 50) = P(10 < e^W < 50) = P(\ln(10) < W > \ln(50))$$
$$= \Phi\left(\frac{\ln(50)-0}{2}\right) - \Phi\left(\frac{\ln(10)-0}{2}\right)$$
$$= \Phi(1.96) - \Phi(1.15) = 0.975002 - 0.874928 = 0.10007$$

b) $P(X < x) = P(e^W < x) = P(W < \ln(x)) = \Phi\left(\frac{\ln(x)-0}{2}\right) = 0.05$

$$\frac{\ln(x)-0}{2} = -1.64 \quad x = e^{-1.64(2)} = 0.0376$$

c) $\mu = E(X) = e^{\theta+\omega^2/2} = e^{0+4/2} = e^2 = 7.389$

$$V(X) = e^{2\theta+\omega^2}(e^{\omega^2}-1) = e^{0+4}(e^4-1) = e^4(e^4-1) = 2926.40$$

4-207. a) Find the values of θ and ω^2 given that E(X) = 50 and V(X) = 4000

$$50 = e^{\theta+\omega^2/2} \qquad 4000 = e^{2\theta+\omega^2}(e^{\omega^2}-1)$$

Let $x = e^{\theta}$ and $y = e^{\omega^2}$ then

$$50 = x\sqrt{y} \text{ and } 4000 = x^2y(y-1) = x^2y^2 - x^2y$$

Square the first equation $x = \frac{50}{\sqrt{y}}$ and substitute into the second equation

$$4000 = \left(\frac{50}{\sqrt{y}}\right)^2 y^2 - \left(\frac{50}{\sqrt{y}}\right)^2 y = 2500(y-1)$$

$$y = 2.6$$

Substitute y back into the first equation and solve for x to obtain $x = \frac{50}{\sqrt{2.6}} = 31$

$$\theta = \ln(31) = 3.43 \text{ and } \omega^2 = \ln(2.6) = 0.96$$

b)

$$P(X < 150) = P(e^W < 150) = P(W < \ln(150)) = \Phi\left(\frac{\ln(150)-3.43}{0.98}\right)$$
$$= \Phi(1.61) = 0.946301$$

4-208. Let X denote the number of fibers visible in a grid cell. Then, X has a Poisson distribution and $\lambda = 100$ fibers per cm^2. Therefore, E(X) = 80,000 fibers per sample = 0.5 fibers per grid cell.

a) $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-0.5}0.5^0}{0!} = 0.3935$.

b) Let W denote the number of grid cells examined until 10 contain fibers. If the number of fibers has a Poisson distribution, then the numbers of fibers in each grid cell are independent. Therefore, W has a negative binomial distribution with p = 0.3935. Consequently, E(W) = 10/0.3935 = 25.41 cells.

c) V(W) = $\frac{10(1-0.3935)}{0.3935^2}$. Therefore, $\sigma_W = 6.25$ cells.

4-209. Let X denote the height of a plant.

a) P(X>2.25) = $P\left(Z > \frac{2.25-2.5}{0.5}\right)$ = P(Z > -0.5) = 1 - P(Z ≤ -0.5) = 0.6915

b) P(2.0 < X < 3.0) = $P\left(\frac{2.0-2.5}{0.5} < Z < \frac{3.0-2.5}{0.5}\right)$ =P(-1 < Z < 1) = 0.683

c.)P(X > x) = 0.90 = $P\left(Z > \frac{x-2.5}{0.5}\right)$ = 0.90 and $\frac{x-2.5}{0.5}$ = -1.28.

Therefore, x = 1.86.

4-210. a) $P(X > 3.5) = \int_{3.5}^{4} (0.5x-1)dx = 0.5\frac{x^2}{2} - x\Big|_{3.5}^{4} = 0.4375$

b) Yes, because the probability of a plant growing to a height of 3.5 centimeters or more without irrigation is small.

4-211. Let X denote the thickness.

a) P(X > 5.5) = $P\left(Z > \frac{5.5-5}{0.2}\right)$ = P(Z > 2.5) = 0. 0062

b) P(4.5 < X < 5.5) = $P\left(\frac{4.5-5}{0.2} < Z < \frac{5.5-5}{0.2}\right)$ = P (-2.5 < Z < 2.5) = 0.9876

Therefore, the proportion that do not meet specifications is 1 − P(4.5 < X < 5.5) = 0.012.

c) If P(X < x) = 0.95, then $P\left(Z > \frac{x-5}{0.2}\right)$ = 0.95. Therefore, $\frac{x-5}{0.2}$ = 1.65 and x = 5.33.

4-212. Let t X denote the dot diameter. If P(0.0014 < X < 0.0026) = 0.9973, then

$$P(\tfrac{0.0014-0.002}{\sigma} < Z < \tfrac{0.0026-0.002}{\sigma}) = P(\tfrac{-0.0006}{\sigma} < Z < \tfrac{0.0006}{\sigma}) = 0.9973.$$

Therefore, $\frac{0.0006}{\sigma}$ = 3 and $\sigma = 0.0002$.

4-213. If P(0.002-x < X < 0.002+x), then P(-x/0.0004 < Z < x/0.0004) = 0.9973. Therefore, x/0.0004 = 3 and x = 0.0012. The specifications are from 0.0008 to 0.0032.

4-214. Let X denote the life.

a) $P(X < 5800) = P(Z < \frac{5800-7000}{600}) = P(Z < -2) = 1 - P(Z \le 2) = 0.023$

b) If P(X > x) = 0.9, then P(Z < $\frac{x-7000}{600}$) = -1.28. Consequently, $\frac{x-7000}{600}$ = -1.28 and x = 6232 hours.

c) If P(X > 10,000) = 0.99, then P(Z > $\frac{10{,}000-\mu}{600}$) = 0.99. Therefore, $\frac{10{,}000-\mu}{600}$ = -2.33 and $\mu = 11{,}398$

d) The probability a product lasts more than 10000 hours is $[P(X > 10000)]^3$, by independence.
If $[P(X > 10000)]^3$ = 0.99, then P(X > 10000) = 0.9967.
Then, P(X > 10000) = $P(Z > \frac{10000-\mu}{600}) = 0.9967$.
Therefore, $\frac{10000-\mu}{600}$ = -2.72 and $\mu = 11{,}632$ hours.

4-215. X is an exponential distribution with E(X) = 7000 hours

a) $P(X < 5800) = \int_0^{5800} \frac{1}{7000} e^{-\frac{x}{7000}} dx = 1 - e^{-\left(\frac{5800}{7000}\right)} = 0.5633$

b) $P(X > x) = \int_x^{\infty} \frac{1}{7000} e^{-\frac{x}{7000}} dx = 0.9$

Therefore, $e^{-\frac{x}{7000}} = 0.9$ and $x = -7000 \ln(0.9) = 737.5$ hours

4-216. Find the values of θand ω^2 given that E(X) = 7000 and σ= 600

$7000 = e^{\theta + \omega^2/2}$ $\quad 600^2 = e^{2\theta + \omega^2}(e^{\omega^2} - 1)$

Let $x = e^{\theta}$ and $y = e^{\omega^2}$ then

$7000 = x\sqrt{y}$ and $600^2 = x^2 y(y-1) = x^2 y^2 - x^2 y$

Square the first equation $7000^2 = x^2 y$ and substitute into the second equation

$600^2 = 7000^2 (y-1)$

$y = 1.0073$

Substitute y into the first equation and solve for x to obtain $x = \frac{7000}{\sqrt{1.0073}} = 6974.6$

$\theta = \ln(6974.6) = 8.850$ and $\omega^2 = \ln(1.0073) = 0.0073$

a)

$$P(X < 5800) = P(e^W < 5800) = P(W < \ln(5800)) = \Phi\left(\frac{\ln(5800) - 8.85}{0.0854}\right)$$

$$= \Phi(-2.16) = 0.015$$

b) $P(X > x) = P(e^W > x) = P(W > \ln(x)) = 1 - \Phi\left(\frac{\ln(x) - 8.85}{0.0854}\right) = 0.9$

$\frac{\ln(x) - 8.85}{0.0854} = -1.28 \quad x = e^{-1.28(0.0854)+8.85} = 6252.20$ hours

4-217.

a) Using the normal approximation to the binomial with n = 50(36)(36) = 64,800and p = 0.0001 we have E(X) = 64800(0.0001) = 6.48

$$P(X \geq 1) \cong P\left(\frac{X-np}{\sqrt{np(1-p)}} \geq \frac{0.5-6.48}{\sqrt{64800(0.0001)(0.9999)}}\right)$$
$$= P(Z > -2.35) = 1-0.0094 = 0.9906$$

b) $$P(X \geq 4) \cong P\left(\frac{X-np}{\sqrt{np(1-p)}} \geq \frac{3.5-6.48}{\sqrt{64800(0.0001)(0.9999)}}\right)$$
$$= P(Z \geq -1.17) = 1-0.1210 = 0.8790$$

4-218.

Using the normal approximation to the binomial with X being the number of people who will be seated. Then X ~Bin(200, 0.9).

a) P(X ≤ 185) $= P\left(\frac{X-np}{\sqrt{np(1-p)}} \leq \frac{185.5-180}{\sqrt{200(0.9)(0.1)}}\right) = P(Z \leq 1.30) = 0.9032$

b)
$P(X < 185)$

$$\approx P(X \leq 184.5) = P\left(\frac{X-np}{\sqrt{np(1-p)}} \geq \frac{184.5-180}{\sqrt{200(0.9)(0.1)}}\right) = P(Z \leq 1.06) = 0.8554$$

c) P(X ≤ 185) ≅ 0.95,
Successively try various values of n. The number of reservations taken could be reduced to 198.

n	Z_o	Probability $P(Z < Z_0)$
190	3.51	0.999776
195	2.39	0.9915758
198	**1.73**	**0.9581849**

4-219.

a)
$$\mu = \frac{\alpha}{\alpha+\beta} = \frac{1}{6} = 0.167$$
$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{5}{6^2 7} = 0.0198$$

b) $P(X > 0.5) = \int_{0.5}^{1} \frac{\Gamma(6)}{\Gamma(1)\Gamma(5)}(1-x)^4 dx = -(1-x)^5 \big|_{0.5}^{1} = \frac{1}{2^5} = 0.03125$

c) $P(X > a) = 0.9 \Rightarrow -(1-x)^5 \big|_{a}^{1} = (1-a)^5 = 0.9 \Rightarrow a = 0.021$

4-220.

Let X denote the amplitude. We know $E(X) = 24$ and $V(X) = 4.1^2$. Determine ω and θ as follows.

$$\mu = E(X) = \exp\left(\theta + \omega^2 / 2\right) = 24$$

$$\sigma^2 = V(X) = \exp\left(2\theta + \omega^2\right)\left(\exp\left(\omega^2\right) - 1\right) = (4.1)^2$$

$$\sigma^2 = \left[\exp\left(\theta + \omega^2 / 2\right)\right]^2\left(\exp\left(\omega^2\right) - 1\right) = \mu^2\left(\exp\left(\omega^2\right) - 1\right) = 16.81$$

Then, $\exp\left(\omega^2\right) - 1 = \dfrac{16.81}{576}$ and $\omega = 0.1696$

$\exp\left(\theta + 0.1696^2 / 2\right) = 24$, then $\theta = 3.16367$

a)

$$P(X > 20) = 1 - P(X \le 20) = 1 - P\left(\exp(W) \le 20\right) = 1 - P\left(W \le \ln(20)\right)$$
$$= 1 - \Phi\left(\frac{\ln(20) - 3.16367}{0.1696}\right) = 1 - 0.16109 = 0.83891$$

b) Find *a* such that $P(X > a) = 0.05$

$$P(X > a) = 1 - P(X \le a) = 1 - P\left(W \le \ln(a)\right) = 1 - \Phi\left(\frac{\ln(a) - 3.16367}{0.1696}\right) = 0.05$$

$\dfrac{\ln(a) - 3.16367}{0.1696} = 1.645$, then $a \cong 3.443$

4.221.

a)

$$P(X > a) = 0.99 \Rightarrow P(Z > \frac{a - 56}{8}) = 0.99 \Rightarrow P(Z < \frac{a - 56}{8}) = 0.01$$
$$\Rightarrow \frac{a - 56}{8} = -2.3260 \Rightarrow a = 37.39$$

b)

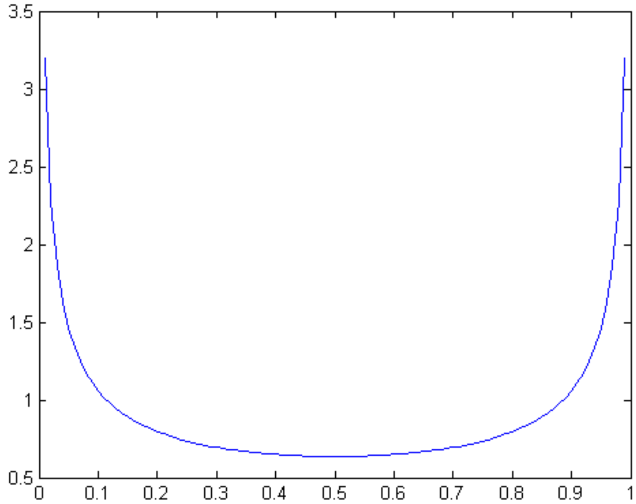
$$P(X > 37.39) = P(Z > \frac{37.39 - \mu}{12}) = 0.01 \Rightarrow \frac{37.39 - \mu}{12} = 2.326 \Rightarrow \mu = 37.39 - 12(2.326) = 9.48$$

c) The subjects can be distinguished well because the means are quite different relative to the standard deviations.

4-222.

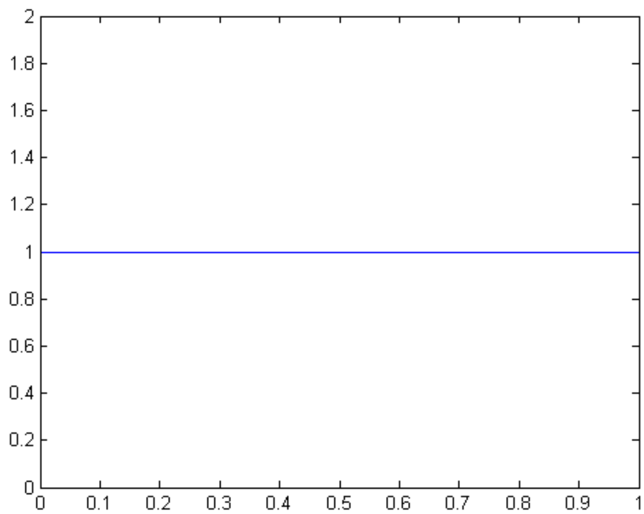
a) $\alpha = \beta = 0.5 < 1$

The function is symmetric and there are two peaks at $x = 0$ and $x = 1$.



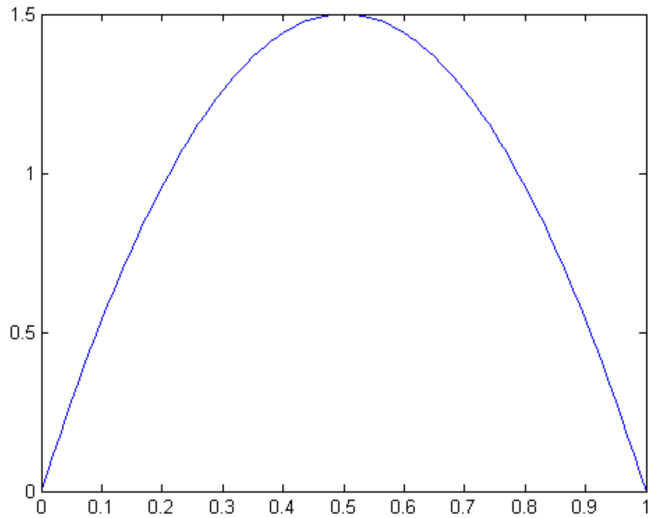
b) $\alpha = \beta = 1$

The function is symmetric and there are no peaks. Actually this probability density function is the same as the standard uniform distribution.



c) $\alpha = \beta = 2 > 1$

The function is symmetric and there is one peak at $x = 0.5$.



4-223.

$$np = 2500$$

$$\sqrt{np(1-p)} = 43.3$$

a) $P(X > 2600) == P(X \geq 2601) = P(Z > \frac{2601 - 0.5 - 2500}{43.3}) = P(Z > 2.321) = 0.01$

b) $P(2400 \leq X \leq 2600) = P(\frac{2400 - 0.5 - 2500}{43.3} < Z < \frac{2600 + 0.5 - 2500}{43.3}) = 0.980$

c)

$$P(X > a) = P(X \geq a+1) = P(Z > \frac{a + 1 - 0.5 - 2500}{43.3}) = 0.05$$

$$\Rightarrow \Phi(\frac{a+1-0.5-2500}{43.3}) = 0.95 \Rightarrow 1.645 = \frac{a+1-0.5-2500}{43.3} \Rightarrow a = 2570.73$$

4-224. Let X denote the time interval between filopodium formation.

a) $P(X > 9) = \exp(-(1/6)\times 9) = 0.223$

b)

$$P(6 < X < 7) = P(X < 7) - P(X \leq 6) = (1 - \exp(-(1/6)\times 7) - (1 - \exp(-(1/6)\times 6) = 0.056$$

c) Find a such that $P(X > a) = 0.9$.

$P(X > a) = \exp(-(1/6)\times a) = 0.9$, then $-\frac{a}{6} = \ln(0.9)$ and $a = 0.632$

4-225.

$$\mu = \frac{\alpha}{\alpha+\beta} = \frac{3.2}{6} = 5.333$$
$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{3.2(2.8)}{6^2 7} = 0.0356$$
$$\mathrm{E}(110X - 60) = 110\mathrm{E}(X) - 60 = -1.333$$
$$V(110X - 60) = 110^2 V(X) = 430.22$$

4-226. Let X denote the survival time of AMI patients.

a) $\delta = 0.25$ (scale parameter); $\beta = 1.16$ (shape parameter)

$$\mu = E(X) = \delta\Gamma\left(1+\frac{1}{\beta}\right) = 0.237$$

$$\sigma^2 = V(X) = \delta^2\Gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2 = 0.042$$

b) $P(X > 1) = 1 - P(X \le 1) = 1 - F(1) = 1 - \left[1 - \exp\left(-\left(\frac{1}{0.25}\right)^{1.16}\right)\right] = 0.0068$

c) Find a such that $P(X > a) = 0.90$

$$P(X > a) = 1 - P(X \le a) = 1 - F(a) = 1 - \left[1 - \exp\left(-\left(\frac{a}{0.25}\right)^{1.16}\right)\right] = 0.90$$

Then, $a = 0.036$.

Mind-Expanding Exercises

4-227. a) P(X > x) implies that there are r - 1 or less counts in an interval of length x. Let Y denote the number of counts in an interval of length x. Then, Y is a Poisson random variable with mena E(Y) = λx. Then, $P(X > x) = P(Y \le r-1) = \sum_{i=0}^{r-1} e^{-\lambda x}\frac{(\lambda x)^i}{i!}$.

b) $P(X \le x) = 1 - \sum_{i=0}^{r-1} e^{-\lambda x}\frac{(\lambda x)^i}{i!}$

c) $f_X(x) = \frac{d}{dx}F_X(x) = \lambda e^{-\lambda x}\sum_{i=0}^{r-1}\frac{(\lambda x)^i}{i!} - e^{-\lambda x}\sum_{i=0}^{r-1}\lambda i\frac{(\lambda x)^i}{i!} = \lambda e^{-\lambda x}\frac{(\lambda x)^{r-1}}{(r-1)!}$

4-228. Let X denote the diameter of the maximum diameter bearing.
Then, P(X > 1.6) = 1 - $P(X \le 1.6)$.

Also, $X \le 1.6$ if and only if all the diameters are less than 1.6. Let Y denote the diameter of a bearing. Then, by independence

$$P(X \le 1.6) = [P(Y \le 1.6)]^{10} = \left[P(Z \le \tfrac{1.6-1.5}{0.025})\right]^{10} = 0.999967^{10} = 0.99967$$

Then, P(X > 1.6) = 0.0033.

4-229. a) Quality loss = $Ek(X-m)^2 = kE(X-m)^2 = k\sigma^2$, by the definition of the variance.

b)

$$\begin{aligned} Quality\ \ loss &= Ek(X-m)^2 = kE(X-\mu+\mu-m)^2 \\ &= kE[(X-\mu)^2 + (\mu-m)^2 + 2(\mu-m)(X-\mu)] \\ &= kE(X-\mu)^2 + k(\mu-m)^2 + 2k(\mu-m)E(X-\mu). \end{aligned}$$

The last term equals zero by the definition of the mean.

Therefore, quality loss = $k\sigma^2 + k(\mu-m)^2$.

4-230.

Let X denote the event that an amplifier fails before 60,000 hours. Let A denote the event that an amplifier mean is 20,000 hours. Then A' is the event that the mean of an amplifier is 50,000 hours. Now, P(E) = P(E|A)P(A) + P(E|A')P(A') and

$$P(E \mid A) = \int_0^{60,000} \tfrac{1}{20,000} e^{-x/20,000} dx = -e^{-x/20,000}\Big|_0^{60,000} = 1 - e^{-3} = 0.9502$$

$$P(E \mid A') = -e^{-x/50,000}\Big|_0^{60,000} = 1 - e^{-6/5} = 0.6988.$$

Therefore, P(E) = 0.9502(0.10) + 0.6988(0.90) = 0.7239

4-231. $P(X < t_1 + t_2 \,|\, X > t_1) = \frac{P(t_1 < X < t_1 + t_2)}{P(X > t_1)}$ from the definition of conditional probability.

Now,

$$P(t_1 < X < t_1 + t_2) = \int_{t_1}^{t_1+t_2} \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_{t_1}^{t_1+t_2} = e^{-\lambda t_1} - e^{-\lambda(t_1+t_2)}$$

$$P(X > t_1) = -e^{-\lambda x}\Big|_{t_1}^{\infty} = e^{-\lambda t_1}$$

Therefore, $P(X < t_1 + t_2 \,|\, X > t_1) = \dfrac{e^{-\lambda t_1}(1 - e^{-\lambda t_2})}{e^{-\lambda t_1}} = 1 - e^{-\lambda t_2} = P(X < t_2)$

4-232.

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = 1, \text{ then } \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$\mu = E(X) = \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1}$$

Suppose $\alpha' = \alpha + 1$, then

$$\mu = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha')\Gamma(\beta)}{\Gamma(\alpha'+\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta}$$

$$\sigma^2 = V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_0^1 x^2 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1}$$

Suppose $\alpha'' = \alpha + 2$, then

$$E(X^2) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha'')\Gamma(\beta)}{\Gamma(\alpha''+\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\alpha(\alpha+1)\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)(\alpha+\beta+1)\Gamma(\alpha+\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

$$\sigma^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha(\alpha+1)(\alpha+\beta)-\alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

We have used the property of the gamma function $\Gamma(r) = (r-1)\Gamma(r-1)$ in this solution.

4-233. $f(x) = \lambda\exp(-\lambda(x-\gamma)), 0 \le \gamma \le x, 0 < \lambda$

a) $\mu = \int_{\gamma}^{\infty} x\lambda\exp(-\lambda(x-\gamma))dx = \gamma + \frac{1}{\lambda}$

$$\sigma^2 = \int_{\gamma}^{\infty} x^2\lambda\exp(-\lambda(x-\gamma))dx - (\gamma+\frac{1}{\lambda})^2 = \gamma^2 + \frac{2\gamma}{\lambda} + \frac{2}{\lambda^2} - (\gamma+\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$$

b) $P(X < \gamma + \frac{1}{\lambda}) = \int_{\gamma}^{\gamma+\frac{1}{\lambda}} \lambda\exp(-\lambda(x-\gamma))dx = -\exp(\lambda(\gamma-x))\,\Big|_{\gamma}^{\gamma+\frac{1}{\lambda}} = 1-\exp(-1) = 0.632$

4-234. a) $1-P(\mu_0-6\sigma < X < \mu_0+6\sigma) = 1-P(-6<Z<6) = 1.97\times10^{-9} = 0.00197\,ppm$

b) $1-P(\mu_0-6\sigma < X < \mu_0+6\sigma) = 1-P(-7.5 < \frac{X-(\mu_0+1.5\sigma)}{\sigma} < 4.5) = 3.4\times10^{-6} = 3.4\,ppm$

c) $1-P(\mu_0-3\sigma < X < \mu_0+3\sigma) = 1-P(-3<Z<3) = .0027 = 2{,}700\,ppm$

d) $1-P(\mu_0-3\sigma < X < \mu_0+3\sigma) = 1-P(-4.5 < \frac{X-(\mu_0+1.5\sigma)}{\sigma} < 1.5)$
$= 0.0668106 = 66{,}810.6\,ppm$

e) If the process is centered six standard deviations away from the specification limits and the process mean shifts even one or two standard deviations there would be minimal product produced outside of specifications. If the process is centered only three standard deviations away from the specifications and the process shifts, there could be a substantial amount of product outside of the specifications.

CHAPTER 5

Section 5-1

5-1. First, f(x,y) $\geq$ 0. Let R denote the range of (X,Y)

Then, $\sum_R f(x,y) = \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} = 1$

a) P(X < 2.5, Y < 3) = f(1.5,2) + f(1,1) = 1/8 + 1/4=3/8
b) P(X < 2.5) = f (1.5, 2) + f (1.5, 3) + f(1,1)= 1/8 + ¼ + 1/4 = 5/8
c) P(Y < 3) = f (1.5, 2) + f(1,1) = 1/8 + 1/4=3/8
d) P(X > 1.8, Y > 4.7) = f (3, 5) = 1/8
e)
E(X) = 1(1/4) + 1.5(3/8) + 2.5(1/4) + 3(1/8) = 1.8125
E(Y) = 1(1/4) + 2(1/8) + 3(1/4) + 4(1/4) + 5(1/8) = 2.875
V(X) = $E(X^2)-[E(X)]^2 = [1^2(1/4) + 1.5^2(3/8) + 2.5^2(1/4) + 3^2(1/8)] - 1.8125^2 = 0.4961$
V(Y) = $E(Y^2)-[E(Y)]^2 = [1^2(1/4) + 2^2(1/8) + 3^2(1/4) + 4^2(1/4) + 5^2(1/8)] - 2.875^2 = 1.8594$

f) Marginal distribution of *X*

x	f(x)
1	¼
1.5	3/8
2.5	¼
3	1/8

g) $f_{Y|1.5}(y) = \dfrac{f_{XY}(1.5, y)}{f_X(1.5)}$ and $f_X(1.5)$ = 3/8. Then,

y	$f_{Y\|1.5}(y)$
2	(1/8)/(3/8)=1/3
3	(1/4)/(3/8)=2/3

h) $f_{X|2}(x) = \dfrac{f_{XY}(x,2)}{f_Y(2)}$ and $f_Y(2)$ = 1/8. Then,

x	$f_{X\|2}(y)$
1.5	(1/8)/(1/8)=1

i) *E*(*Y*|*X* = 1.5) = 2(1/3) + 3(2/3) =2 1/3
j) Because $f_{Y|1.5}(y) \neq f_Y(y)$, X and Y are not independent.

5-2. Let R denote the range of (X,Y). Because

$\sum_R f(x,y) = c(2+3+4+3+4+5+4+5+6) = 1$, 36c = 1, and c = 1/36

a) $P(X = 1, Y < 4) = f_{XY}(1,1) + f_{XY}(1,2) + f_{XY}(1,3) = \frac{1}{36}(2+3+4) = 1/4$
b) *P*(*X* = 1) is the same as part (a) = 1/4
c) $P(Y = 2) = f_{XY}(1,2) + f_{XY}(2,2) + f_{XY}(3,2) = \frac{1}{36}(3+4+5) = 1/3$
d) $P(X < 2, Y < 2) = f_{XY}(1,1) = \frac{1}{36}(2) = 1/18$

e)

$$E(X) = 1[f_{XY}(1,1) + f_{XY}(1,2) + f_{XY}(1,3)] + 2[f_{XY}(2,1) + f_{XY}(2,2) + f_{XY}(2,3)]$$
$$+ \ 3[f_{XY}(3,1) + f_{XY}(3,2) + f_{XY}(3,3)]$$
$$= \left(1\times\tfrac{9}{36}\right) + \left(2\times\tfrac{12}{36}\right) + \left(3\times\tfrac{15}{36}\right) = 13/6 = 2.167$$

$$V(X) = (1-\tfrac{13}{6})^2\ \tfrac{9}{36} + (2-\tfrac{13}{6})^2\ \tfrac{12}{36} + (3-\tfrac{13}{6})^2\ \tfrac{15}{36} = 0.639$$

$$E(Y) = 2.167$$

$$V(Y) = 0.639$$

f) Marginal distribution of *X*

x	$f_X(x) = f_{XY}(x,1) + f_{XY}(x,2) + f_{XY}(x,3)$
1	1/4
2	1/3
3	5/12

g) $f_{Y|X}(y) = \dfrac{f_{XY}(1,y)}{f_X(1)}$

y	$f_{Y\|X}(y)$
1	(2/36)/(1/4)=2/9
2	(3/36)/(1/4)=1/3
3	(4/36)/(1/4)=4/9

h) $f_{X|Y}(x) = \dfrac{f_{XY}(x,2)}{f_Y(2)}$ and $f_Y(2) = f_{XY}(1,2) + f_{XY}(2,2) + f_{XY}(3,2) = \tfrac{12}{36} = 1/3$

x	$f_{X\|Y}(x)$
1	(3/36)/(1/3)=1/4
2	(4/36)/(1/3)=1/3
3	(5/36)/(1/3)=5/12

i) $E(Y|X = 1) = 1(2/9) + 2(1/3) + 3(4/9) = 20/9$

j) Since $f_{XY}(x,y) \neq f_X(x)f_Y(y)$, X and Y are not independent.

5-3. $f(x,y) \geq 0$ *and* $\sum_R f(x,y) = 1$

a) $P(X < 0.5, Y < 1.5) = f_{XY}(-1,-2) + f_{XY}(-0.5,-1) = \tfrac{1}{8} + \tfrac{1}{4} = \tfrac{3}{8}$

b) $P(X < 0.5) = f_{XY}(-1,-2) + f_{XY}(-0.5,-1) = \tfrac{3}{8}$

c) $P(Y < 1.5) = f_{XY}(-1,-2) + f_{XY}(-0.5,-1) + f_{XY}(0.5,1) = \tfrac{7}{8}$

d) $P(X > 0.25, Y < 4.5) = f_{XY}(0.5,1) + f_{XY}(1,2) = \tfrac{5}{8}$

e)

$$E(X) = -1(\tfrac{1}{8}) - 0.5(\tfrac{1}{4}) + 0.5(\tfrac{1}{2}) + 1(\tfrac{1}{8}) = \tfrac{1}{8}$$

$$E(Y) = -2(\tfrac{1}{8}) - 1(\tfrac{1}{4}) + 1(\tfrac{1}{2}) + 2(\tfrac{1}{8}) = \tfrac{1}{4}$$

$V(X) = (-1 - 1/8)^2(1/8) + (-0.5 - 1/8)^2(1/4) + (0.5 - 1/8)^2(1/2) + (1 - 1/8)^2(1/8) = 0.4219$

$V(Y) = (-2 - 1/4)^2(1/8) + (-1 - 1/4)^2(1/4) + (1 - 1/4)^2(1/2) + (2 - 1/4)^2(1/8) = 1.6875$

f) Marginal distribution of *X*

x	$f_X(x)$
-1	1/8
-0.5	¼
0.5	½
1	1/8

g) $f_{Y|X}(y) = \dfrac{f_{XY}(1, y)}{f_X(1)}$

y	$f_{Y\|X}(y)$
2	1/8/(1/8)=1

h) $f_{X|Y}(x) = \dfrac{f_{XY}(x,1)}{f_Y(1)}$

x	$f_{X\|Y}(x)$
0.5	½/(1/2)=1

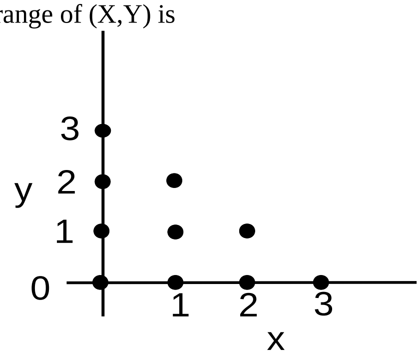
i) $E(X|Y = 1) = 0.5$

j) No, *X* and *Y* are not independent.

5-4. Because X and Y denote the number of printers in each category,

$$X \geq 0,\ Y \geq 0 \ \ and\ \ X + Y = 4$$

5-5. a) The range of (X,Y) is



x,y	$f_{xy}(x,y)$
0,0	0.857375
0,1	0.1083
0,2	0.00456
0,3	0.000064
1,0	0.027075

1,1	0.00228
1,2	0.000048
2,0	0.000285
2,1	0.000012
3,0	0.000001

b)

x	$f_x(x)$
0	0.970299
1	0.029403
2	0.000297
3	0.000001

c) E(X) = 0(0.970299) + 1(0.029403) + 2(0.000297) + 3(0.000001) = 0.03
or np = 3(0.01) = 0.03

d) $f_{Y|1}(y) = \dfrac{f_{XY}(1, y)}{f_X(1)}$, $f_x(1) = 0.029403$

y	$f_{Y\|1}(x)$
0	0.920824
1	0.077543
2	0.001632

e) E(Y|X = 1) = 0(.920824) + 1(0.077543) + 2(0.001632) = 0.080807
g) No, X and Y are not independent because, for example, $f_Y(0) \neq f_{Y|1}(0)$.

5-6.

a) The range of (X,Y) is $X \geq 0,\ Y \geq 0\ and\ X + Y \leq 4$. Here X is the number of pages with moderate graphic content and Y is the number of pages with high graphic output among a sample of 4 pages.

The following table is for sampling without replacement. Students would have to extend the hypergeometric distribution to the case of three classes (low, moderate, and high).

For example, *P*(*X* = 1, *Y* = 2) is calculated as

$$P(X = 1, Y = 2) = \frac{\binom{60}{1}\binom{30}{1}\binom{10}{2}}{\binom{100}{4}} = \frac{60(30)(45)}{\frac{100(99)(98)(97)}{24}} = 0.02066$$

	x=0	x=1	x=2	x=3	x=4
y=4	5.35×10^{-05}	0	0	0	0
y=3	0.00184	0.00092	0	0	0
y=2	0.02031	0.02066	0.00499	0	0
y=1	0.08727	0.13542	0.06656	0.01035	0
y=0	0.12436	0.26181	0.19635	0.06212	0.00699

b)

x=0 x=1 x=2 x=3 x=4

f(x)	0.2338	0.4188	0.2679	0.0725	0.0070

c) E(X)=

$$\sum_{0}^{4} x_i f(x_i) = 0(0.2338) + 1(0.4188) + 2(0.2679) + 3(0.7248) = 4(0.0070) = 1.2$$

d) $f_{Y|3}(y) = \dfrac{f_{XY}(3,y)}{f_X(3)}$, $f_x(3) = 0.0725$

y	$f_{Y\|3}(y)$
0	0.857
1	0.143
2	0
3	0
4	0

e) $E(Y|X = 3) = 0(0.857) + 1(0.143) = 0.143$
f) $V(Y|X = 3) = 0^2(0.857) + 1^2(0.143) - 0.143^2 = 0.123$
g) No, *X* and *Y* are not independent

5-7. a) The range of (X,Y) is $X \geq 0,\ Y \geq 0\ and\ X + Y \leq 4$.

Here *X* and *Y* denote the number of defective items found with inspection devices 1 and 2, respectively.

	x=0	x=1	x=2	x=3	x=4
y=0	1.94×10^{-19}	1.10×10^{-16}	2.35×10^{-14}	2.22×10^{-12}	7.88×10^{-11}
y=1	2.59×10^{-16}	1.47×10^{-13}	3.12×10^{-11}	2.95×10^{-9}	1.05×10^{-7}
y=2	1.29×10^{-13}	7.31×10^{-11}	1.56×10^{-8}	1.47×10^{-6}	5.22×10^{-5}
y=3	2.86×10^{-11}	1.62×10^{-8}	3.45×10^{-6}	3.26×10^{-4}	0.0116
y=4	2.37×10^{-9}	1.35×10^{-6}	2.86×10^{-4}	0.0271	0.961

$$f(x,y) = \left[\binom{4}{x}(0.993)^x(0.007)^{4-x}\right]\left[\binom{4}{y}(0.997)^y(0.003)^{4-y}\right]$$

For *x* = 1, 2, 3, 4 and *y* = 1, 2, 3,4

b)

$$f(x,y) = \left[\binom{4}{x}(0.993)^x(0.007)^{4-x}\right] \text{for } x = 1,2,3,4$$

	x=0	x=1	x=2	x=3	x=4
f(x)	2.40×10^{-9}	1.36×10^{-6}	2.899×10^{-4}	0.0274	0.972

c) Because X has a binomial distribution $E(X) = n(p) = 4(0.993) = 3.972$

d) $f_{Y|2}(y) = \dfrac{f_{XY}(2,y)}{f_X(2)} = f(y)$, $f_x(2) = 2.899 \times 10^{-4}$

y	$f_{Y\|1}(y)=f(y)$
0	8.1×10^{-11}

1	1.08×10^{-7}
2	5.37×10^{-5}
3	0.0119
4	0.988

e) E(Y|X = 2) = E(Y) = $n(p)$ = 4(0.997) = 3.988
f) V(Y|X = 2) = V(Y) = $n(p)(1-p)$ = 4(0.997)(0.003) = 0.0120
g) Yes, *X* and *Y* are independent.

5-8. a) $P(X=2) = f_{XYZ}(2,1,1) + f_{XYZ}(2,1,2) + f_{XYZ}(2,2,1) + f_{XYZ}(2,2,2) = 0.5$

b) $P(X=1, Y=2) = f_{XYZ}(1,2,1) + f_{XYZ}(1,2,2) = 0.35$

c) c) $P(Z<1.5) = f_{XYZ}(1,1,1) + f_{XYZ}(1,2,1) + f_{XYZ}(2,1,1) + f_{XYZ}(2,2,1) = 0.5$

d)
$$P(X=1 \text{ or } Z=2) = P(X=1) + P(Z=2) - P(X=1, Z=2) = 0.5 + 0.5 - 0.3 = 0.7$$

e) E(X) = 1(0.5) + 2(0.5) = 1.5

f) $P(X=1 \mid Y=1) = \dfrac{P(X=1, Y=1)}{P(Y=1)} = \dfrac{0.05+0.10}{0.15+0.2+0.1+0.05} = 0.3$

g) $P(X=1, Y=1 \mid Z=2) == \dfrac{P(X=1, Y=1, Z=2)}{P(Z=2)} = \dfrac{0.1}{0.1+0.2+0.15+0.05} = 0.2$

h) $P(X=1 \mid Y=1, Z=2) = \dfrac{P(X=1, Y=1, Z=2)}{P(Y=1, Z=2)} = \dfrac{0.10}{0.10+0.15} = 0.4$

i) $f_{X|YZ}(x) = \dfrac{f_{XYZ}(x,1,2)}{f_{YZ}(1,2)}$ and $f_{YZ}(1,2) = f_{XYZ}(1,1,2) + f_{XYZ}(2,1,2) = 0.25$

x	$f_{X\|YZ}(x)$
1	0.10/0.25=0.4
2	0.15/0.25=0.6

5-9. Number of students:
Electrical 24
Industrial 4
Mechanical 12
X and Y = numbers of industrial and mechanical students in the sample, respectively

(a)

$$P(X=x, Y=y) = \frac{\binom{4}{x}\binom{12}{y}\binom{24}{4-x-y}}{\binom{40}{4}} \text{ for x + y} \le 4$$

x	y	f(x,y)
0	0	0.116271
0	1	0.265762
0	2	0.199322
0	3	0.057774
0	4	0.005416
1	0	0.088587
1	1	0.144961
1	2	0.069329
1	3	0.009629
2	0	0.01812
2	1	0.018908
2	2	0.004333
3	0	0.00105
3	1	0.000525
4	0	1.09E-05

b) $f_X(x) = P(X = x) = \sum_{\{y|x+y\le 4\}} f_{XY}(x, y)$

x	f(x)
0	0.644545
1	0.312507
2	0.041361
3	0.001576
4	1.09E-05

c) $E(X) = \sum x f_X(x)$ =0(0.6445) + 1(0.3125) + 2(0.0414) + 3(0.0016) + 4(1.09E-5) = 0.4

d) f(y|X = 3) = P(Y = y, X = 3)/P(X = 3)

$P(Y = 0, X = 3) = C^4_3\, C^{12}_0 C^{24}_1 / C^{40}_4$
$P(Y = 1, X=3) = C^4_3\, C^{12}_1 C^{24}_0 / C^{40}_4$

$P(X = 3) = C^{36}_1\, C^4_3 / C^{40}_4$, from the hypergeometric distribution with N = 40, n = 4, k = 4, x = 3

Therefore

$f(0|X=3) = [C^{24}_1\, C^4_3 / C^{40}_4]/[C^{36}_1\, C^4_3 / C^{40}_4] = C^{24}_1 / C^{36}_1 = 2/3$

$f(1|X=3) = [C^{12}_1\, C^4_3 / C^{40}_4]/[C^{36}_1\, C^4_3 / C^{40}_4] = C^{12}_1 / C^{36}_1 = 1/3$

y	x	$f_{Y\|3}(y)$
0	3	2/3
1	3	1/3

e) E(Y|X = 3) = 0(0.6667) + 1(0.3333)=0.3333
f) V(Y|X = 3) = $(0 - 0.3333)^2(0.6667) + (1-0.3333)^2(0.3333) = 0.0741$
g) $f_X(0) = 0.2511$, $f_Y(0) = 0.1555$, $f_X(0)f_Y(0) = 0.039046 \neq f_{XY}(0,0) = 0.1296$
X and Y are not independent.

5-10. (a) P(X < 5) = 0.44 + 0.04 = 0.48
(b) E(X) = 0.43(23) + 0.44(4.2) + 0.04(11.4) + 0.05(130) + 0.04(0) = 18.694
(c) $P_{X|Y=0}(X)$ = P(X = x,Y = 0)/P(Y = 0) = 0.04/0.08 = 0.5 for x = 0 and 11.4
(d) P(X < 6|Y = 0) = P(X = 0|Y = 0) = 0.5
(e) E(X|Y = 0)=11.4(0.5) + 0(0.5) = 5.7

5-11. (a) $f_{XYZ}(x,y,z)$

$f_{XYZ}(x,y,z)$	Selects(X)	Updates(Y)	Inserts(Z)
0.43	23	11	12
0.44	4.2	3	1
0.04	11.4	0	0
0.05	130	120	0
0.04	0	0	0

(b) $P_{XY|Z=0}$

$P_{XY\|Z=0}(x,y)$	Selects(X)	Updates(Y)	Inserts(Z)
4/13 = 0.3077	11.4	0	0
5/13 = 0.3846	130	120	0
4/13= 0.3077	0	0	0

(c) P(X<6, Y<6|Z = 0) = P(X =0, Y = 0) = 0.3077

(d) E(X|Y = 0,Z = 0) = 0.5(11.4) + 0.5(0) = 5.7 where this conditional distribuition for X was determined in the previous exercise

5-12. Let X, Y, and Z denote the number of bits with high, moderate, and low distortion. Then, the joint distribution of X, Y, and Z is multinomial with n = 3 and
$p_1 = 0.01,\ \ p_2 = 0.04,\ \ and\ \ p_3 = 0.95$

a)
$$P(X=2,Y=1) = P(X=2,Y=1,Z=0)$$
$$= \frac{3!}{2!1!0!}0.01^2 0.04^1 0.95^0 = 1.2\times 10^{-5}$$

b) $P(X=0,Y=0,Z=3) = \dfrac{3!}{0!0!3!}0.01^0 0.04^0 0.95^3 = 0.8574$

c) X has a binomial distribution with n = 3 and p = 0.01. Then, E(X) = 3(0.01) = 0.03 and V(X) = 3(0.01)(0.99) = 0.0297

d) First determine $P(X \mid Y=2)$

$$P(Y=2)=P(X=1,Y=2,Z=0)+P(X=0,Y=2,Z=1)$$
$$=\frac{3!}{1!2!0!}0.01(0.04)^2 0.95^0+\frac{3!}{0!2!1!}0.01^0(0.04)^2 0.95^1=0.0046$$
$$P(X=0\,|\,Y=2)=\frac{P(X=0,Y=2)}{P(Y=2)}=\left(\frac{3!}{0!2!1!}0.01^0 0.04^2 0.95^1\right)\Big/0.004608=0.98958$$

$$P(X=1\,|\,Y=2)=\frac{P(X=1,Y=2)}{P(Y=2)}=\left(\frac{3!}{1!2!1!}0.01^1 0.04^2 0.95^0\right)\Big/0.004608=0.01042$$
$$E(X\,|\,Y=2)=0(0.98958)+1(0.01042)=0.01042$$

$$V(X\,|\,Y=2)=E(X^2)-(E(X))^2=0.01042-(0.01042)^2=0.01031$$

5-13. Determine c such that $c\int_0^3\int_0^3 xydxdy=c\int_0^3 y\left.\frac{x^2}{2}\right|_0^3 dy=c(4.5\left.\frac{y^2}{2}\right|_0^3)=\frac{81}{4}c.$

Therefore, c = 4/81.

a) $P(X<2,Y<3)=\frac{4}{81}\int_0^3\int_0^2 xydxdy=\frac{4}{81}(2)\int_0^3 ydy=\frac{4}{81}(2)(\frac{9}{2})=0.4444$

b) P(X < 2.5) = P(X < 2.5, Y < 3) because the range of Y is from 0 to 3.

$$P(X<2.5,Y<3)=\frac{4}{81}\int_0^3\int_0^{2.5} xydxdy=\frac{4}{81}(3.125)\int_0^3 ydy=\frac{4}{81}(3.125)\frac{9}{2}=0.6944$$

c) $P(1<Y<2.5)=\frac{4}{81}\int_1^{2.5}\int_0^3 xydxdy=\frac{4}{81}(4.5)\int_1^{2.5} ydy=\frac{18}{81}\left.\frac{y^2}{2}\right|_1^{2.5}=0.5833$

d) $P(X>1.8,1<Y<2.5)=\frac{4}{81}\int_1^{2.5}\int_{1.8}^3 xydxdy=\frac{4}{81}(2.88)\int_1^{2.5} ydy=\frac{4}{81}(2.88)\frac{(2.5^2-1)}{2}=0.3733$

e) $E(X)=\frac{4}{81}\int_0^3\int_0^3 x^2 ydxdy=\frac{4}{81}\int_0^3 9ydy=\frac{4}{9}\left.\frac{y^2}{2}\right|_0^3=2$

f) $P(X<0,Y<4)=\frac{4}{81}\int_0^4\int_0^0 xydxdy=0\int_0^4 ydy=0$

g) $f_X(x)=\int_0^3 f_{XY}(x,y)dy=x\frac{4}{81}\int_0^3 ydy=\frac{4}{81}x(4.5)=\frac{2x}{9}\quad for\quad 0<x<3.$

h) $f_{Y|1.5}(y)=\frac{f_{XY}(1.5,y)}{f_X(1.5)}=\frac{\frac{4}{81}y(1.5)}{\frac{2}{9}(1.5)}=\frac{2}{9}y$ for $0 < y < 3$.

i) E(Y|X=1.5) = $\int_0^3 y\left(\frac{2}{9}y\right)dy=\frac{2}{9}\int_0^3 y^2 dy=\left.\frac{2y^3}{27}\right|_0^3=2$

j) $P(Y < 2 \mid X = 1.5) = f_{Y|1.5}(y) = \int_0^2 \frac{2}{9} y\,dy = \frac{1}{9} y^2 \Big|_0^2 = \frac{4}{9} - 0 = \frac{4}{9}$

k) $f_{X|2}(x) = \frac{f_{XY}(x,2)}{f_Y(2)} = \frac{\frac{4}{81}x(2)}{\frac{2}{9}(2)} = \frac{2}{9}x$ for 0 < x < 3.

5-14.

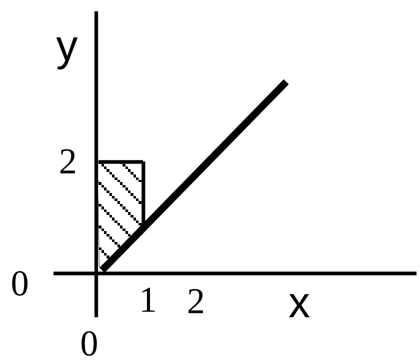
$$c\int_0^3 \int_x^{x+2} (x+y)\,dy\,dx = \int_0^3 xy + \frac{y^2}{2}\Big|_x^{x+2} dx$$

$$= \int_0^3 \left[x(x+2) + \frac{(x+2)^2}{2} - x^2 - \frac{x^2}{2}\right] dx$$

$$= c\int_0^3 (4x+2)\,dx = \left[2x^2 + 2x\right]_0^3 = 24c$$

Therefore, c = 1/24.

a) P(X < 1, Y < 2) equals the integral of $f_{XY}(x,y)$ over the following region.

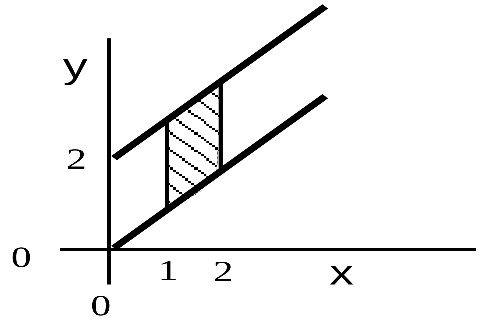


Then,

$$P(X < 1, Y < 2) = \frac{1}{24}\int_0^1 \int_x^2 (x+y)\,dy\,dx = \frac{1}{24}\int_0^1 xy + \frac{y^2}{2}\Big|_x^2 dx = \frac{1}{24}\int_0^3 2x + 2 - \frac{3x^2}{2} dx =$$

$$\frac{1}{24}\left[x^2 + 2x - \frac{x^3}{2}\Big|_0^1\right] = 0.10417$$

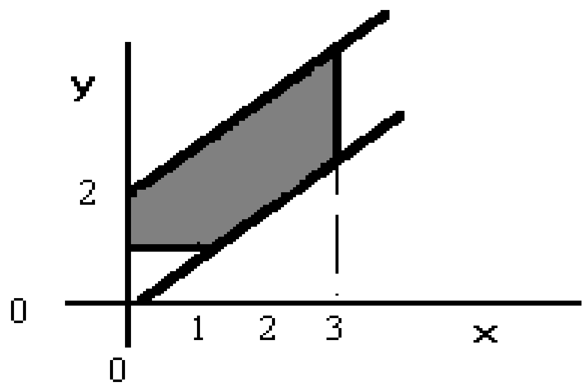
b) P(1 < X < 2) equals the integral of $f_{XY}(x,y)$ over the following region



$$P(1 < X < 2) = \frac{1}{24}\int_1^2 \int_x^{x+2} (x+y)dydx = \frac{1}{24}\int_1^2 xy + \tfrac{y^2}{2}\Big|_x^{x+2} dx$$

$$= \frac{1}{24}\int_0^3 (4x+2)dx = \frac{1}{24}\left[2x^2 + 2x\Big|_1^2\right] = \tfrac{1}{6}.$$

c) P(Y > 1) is the integral of $f_{XY}(x, y)$ over the following region.

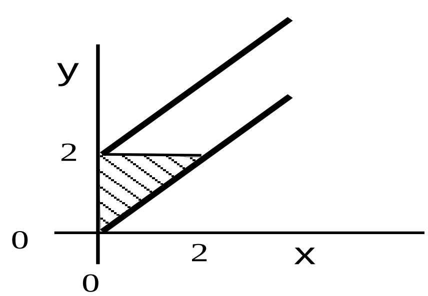


$$P(Y > 1) = 1 - P(Y \le 1) = 1 - \tfrac{1}{24}\int_0^1 \int_x^1 (x+y)dydx = 1 - \frac{1}{24}\int_0^1 (xy + \frac{y^2}{2})\Big|_x^1$$

$$= 1 - \frac{1}{24}\int_0^1 x + \frac{1}{2} - \frac{3}{2}x^2 dx = 1 - \frac{1}{24}\left(\frac{x^2}{2} + \frac{1}{2} - \frac{1}{2}x^3\right)\Bigg|_0^1$$

$$= 1 - 0.02083 = 0.9792$$

d) P(X < 2, Y < 2) is the integral of $f_{XY}(x,y)$ over the following region.



$$E(X)=\frac{1}{24}\int_0^3\int_x^{x+2} x(x+y)dydx=\frac{1}{24}\int_0^3 x^2y+\left.\tfrac{xy^2}{2}\right|_x^{x+2}dx$$
$$=\frac{1}{24}\int_0^3(4x^2+2x)dx=\frac{1}{24}\left[\frac{4x^3}{3}+\left.x^2\right|_0^3\right]=\frac{15}{8}$$

e)

$$E(X)=\frac{1}{24}\int_0^3\int_x^{x+2} x(x+y)dydx=\frac{1}{24}\int_0^3 x^2y+\left.\tfrac{xy^2}{2}\right|_x^{x+2}dx$$
$$=\frac{1}{24}\int_0^3(4x^2+2x)dx=\frac{1}{24}\left[\frac{4x^3}{3}+\left.x^2\right|_0^3\right]=\frac{15}{8}$$

f)

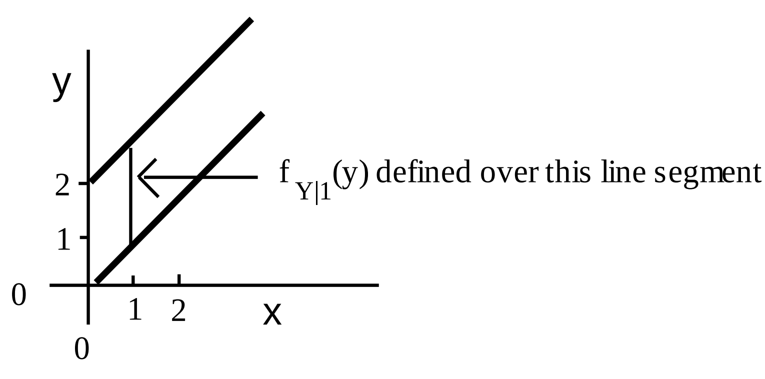
$$V(X)=\frac{1}{24}\int_0^3\int_x^{x+2} x^2(x+y)dydx-\left(\frac{15}{8}\right)^2=\frac{1}{24}\int_0^3 x^3y+\left.\tfrac{x^2y^2}{2}\right|_x^{x+2}dx-\left(\frac{15}{8}\right)^2$$
$$=\frac{1}{24}\int_0^3(3x^3+4x^2+4x-\frac{x^4}{4})dx-\left(\frac{15}{8}\right)^2$$
$$=\frac{1}{24}\left[\frac{3x^4}{4}+\frac{4x^3}{3}+2x^2-\left.\frac{x^5}{20}\right|_0^3\right]-\left(\frac{15}{8}\right)^2=\frac{31707}{320}$$

g) $f_X(x)$ is the integral of $f_{XY}(x,y)$ over the interval from x to x+2. That is,

$$f_X(x)=\frac{1}{24}\int_x^{x+2}(x+y)dy=\frac{1}{24}\left[xy+\left.\tfrac{y^2}{2}\right|_x^{x+2}\right]=\frac{x}{6}+\frac{1}{12}\quad \text{for } 0 < x < 3.$$

h) $f_{Y|1}(y)=\frac{f_{XY}(1,y)}{f_X(1)}=\frac{\frac{1}{24}(1+y)}{\frac{1}{6}+\frac{1}{12}}=\frac{1+y}{6}$ for $1 < y < 3$.

See the following graph



i) $E(Y|X = 1) = \int_1^3 y\left(\frac{1+y}{6}\right)dy=\frac{1}{6}\int_1^3(y+y^2)dy=\frac{1}{6}\left.\left(\frac{y^2}{2}+\frac{y^3}{3}\right)\right|_1^3=2.111$

j) $P(Y>2 \mid X=1) = \int_2^3 \left(\frac{1+y}{6}\right)dy = \frac{1}{6}\int_2^3 (1+y)dy = \frac{1}{6}\left(y+\frac{y^2}{2}\right)\Big|_2^3 = 0.5833$

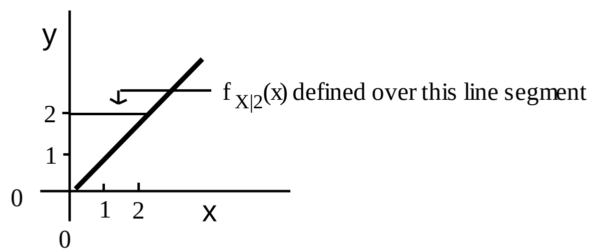
k) $f_{X|2}(x) = \frac{f_{XY}(x,2)}{f_Y(2)}$. Here $f_Y(y)$ is determined by integrating over *x*. There are three regions of integration.

For $0<y\le 2$ the integration is from 0 to *y*.

For $2<y\le 3$ the integration is from *y*-2 to *y*.

For $3<y<5$ the integration is from *y* to 3. Because the condition is *y* = 2, only the first integration is needed.

$$f_Y(y) = \frac{1}{24}\int_0^y (x+y)dx = \frac{1}{24}\left[\frac{x^2}{2}+xy\Big|_0^y\right] = \frac{y^2}{16} \quad \text{for } 0<y\le 2.$$



Therefore, $f_Y(2)=1/4$ and $f_{X|2}(x) = \frac{\frac{1}{24}(x+2)}{1/4} = \frac{x+2}{6}$ for 0 < x < 3

5-15. $c\int_0^3\int_0^x xy\,dy\,dx = c\int_0^3 x\frac{y^2}{2}\Big|_0^x dx = c\int_0^3 \frac{x^3}{2}dx = c\frac{x^4}{8}\Big|_0^3 = \frac{81}{8}c.$ Therefore, *c* = 8/81

a) P(X<1,Y<2)= $\frac{8}{81}\int_0^1\int_0^x xy\,dy\,dx = \frac{8}{81}\int_0^1 \frac{x^3}{2}dx = \frac{8}{81}\left(\frac{1}{8}\right) = \frac{1}{81}$

b) P(1<X<2) = $\frac{8}{81}\int_1^2\int_0^x xy\,dy\,dx = \frac{8}{81}\int_1^2 x\frac{x^2}{2}dx = \left(\frac{8}{81}\right)\frac{x^4}{8}\Big|_1^2 = \left(\frac{8}{81}\right)\frac{(2^4-1)}{8} = \frac{5}{27}$

c)

$$P(Y>1) = \frac{8}{81}\int_1^3\int_1^x xy\,dy\,dx = \frac{8}{81}\int_1^3 x\left(\frac{x^2-1}{2}\right)dx = \frac{8}{81}\int_1^3 \frac{x^3}{2} - \frac{x}{2}dx = \frac{8}{81}\left(\frac{x^4}{8}-\frac{x^2}{4}\right)\Big|_1^3$$

$$= \frac{8}{81}\left[\left(\frac{3^4}{8}-\frac{3^2}{4}\right)-\left(\frac{1^4}{8}-\frac{1^2}{4}\right)\right] = \frac{64}{81} = 0.7901$$

d) P(X<2,Y<2) = $\frac{8}{81}\int_0^2\int_0^x xy\,dy\,dx = \frac{8}{81}\int_0^2 \frac{x^3}{2}dx = \frac{8}{81}\left(\frac{2^4}{8}\right) = \frac{16}{81}.$

e)

$$E(X) = \frac{8}{81}\int_0^3\int_0^x x(xy)dydx = \frac{8}{81}\int_0^3\int_0^x x^2 ydydx = \frac{8}{81}\int_0^3 \frac{x^2}{2}x^2dx = \frac{8}{81}\int_0^3 \frac{x^4}{2}dx$$

$$= \left(\frac{8}{81}\right)\left(\frac{3^5}{10}\right) = \frac{12}{5}$$

f)

$$E(Y) = \frac{8}{81}\int_0^3\int_0^x y(xy)dydx = \frac{8}{81}\int_0^3\int_0^x xy^2dydx = \frac{8}{81}\int_0^3 x\frac{x^3}{3}dx$$

$$= \frac{8}{81}\int_0^3 \frac{x^4}{3}dx = \left(\frac{8}{81}\right)\left(\frac{3^5}{15}\right) = \frac{8}{5}$$

g) $f(x) = \frac{8}{81}\int_0^x xydy = \frac{4x^3}{81} \qquad 0 < x < 3$

h) $f_{Y|x=1}(y) = \frac{f(1,y)}{f(1)} = \frac{\frac{8}{81}(1)y}{\frac{4(1)^3}{81}} = 2y \qquad 0 < y < 1$

i) $E(Y \mid X = 1) = \int_0^1 2ydy = y^2\Big|_0^1 = 1$

j) $P(Y > 2 \mid X = 1) = 0$ this isn't possible since the values of y are 0< $y < x$.

k) $f(y) = \frac{8}{81}\int_y^3 xydx = \frac{4}{9}y - \frac{4}{81}y^3, 0 < y < 3$, Therefore

$$f_{X|Y=2}(x) = \frac{f_{XY}(x,2)}{f_Y(2)} = \frac{\frac{8}{81}x(2)}{\frac{4}{9}(2) - \frac{4}{81}(8)} = \frac{2x}{5} \qquad 2 < x < 3$$

5-16. Solve for c

$$c\int_0^\infty\int_0^x e^{-2x-3y}dydx = \frac{c}{3}\int_0^\infty e^{-2x}\left(1-e^{-3x}\right)dx = \frac{c}{3}\int_0^\infty e^{-2x} - e^{-5x}dx =$$

$$\frac{c}{3}\left(\frac{1}{2}-\frac{1}{5}\right) = \frac{1}{10}c. \quad c = 10$$

a)

$$P(X < 1, Y < 2) = 10\int_0^1\int_0^x e^{-2x-3y}dydx = \frac{10}{3}\int_0^1 e^{-2x}(1-e^{-3x})dy = \frac{10}{3}\int_0^1 e^{-2x} - e^{-5x}dy$$

$$= \frac{10}{3}\left(\frac{e^{-5x}}{5} - \frac{e^{-2x}}{2}\right)\Bigg|_0^1 = 0.77893$$

b)

$$P(1 < X < 2) = 10\int_1^2\int_0^x e^{-2x-3y}\,dy\,dx = \frac{10}{3}\int_1^2 e^{-2x} - e^{-5x}\,dx$$

$$= \frac{10}{3}\left(\frac{e^{-5x}}{5} - \frac{e^{-2x}}{2}\right)\Bigg|_1^2 = 0.19057$$

c)

$$P(Y > 3) = 10\int_3^\infty\int_3^x e^{-2x-3y}\,dy\,dx = \frac{10}{3}\int_3^{1\infty} e^{-2x}(e^{-9} - e^{-3x})\,dy$$

$$= \frac{10}{3}\left(\frac{e^{-5x}}{5} - \frac{e^{-9}e^{-2x}}{2}\right)\Bigg|_3^\infty = 3.059x10^{-7}$$

d)

$$P(X < 2, Y < 2) = 10\int_0^2\int_0^x e^{-2x-3y}\,dy\,dx = \frac{10}{3}\int_0^2 e^{-2x}(1 - e^{-3x})\,dx = \frac{10}{3}\left(\frac{e^{-10}}{5} - \frac{e^{-4}}{2}\right)\Bigg|_0^2$$

$$= 0.9695$$

e) E(X) $= 10\int_0^\infty\int_0^x xe^{-2x-3y}\,dy\,dx = \frac{7}{10}$

f) E(Y) $= 10\int_0^\infty\int_0^x ye^{-2x-3y}\,dy\,dx = \frac{1}{5}$

g) $f(x) = 10\int_0^x e^{-2x-3y}\,dy = \frac{10e^{-2z}}{3}(1 - e^{-3x}) = \frac{10}{3}(e^{-2x} - e^{-5x})$ for $0 < x$

h) $f_{Y\setminus X=1}(y) = \frac{f_{X,Y}(1,y)}{f_X(1)} = \frac{10e^{-2-3y}}{\frac{10}{3}(e^{-2} - e^{-5})} = 3.157e^{-3y}$ $0 < y < 1$

i) $E(Y|X=1) = 3.157\int_0^1 ye^{-3y}\,dy = 0.2809$

j) $f_{X|Y=2}(x) = \frac{f_{X,Y}(x,2)}{f_Y(2)} = \frac{10e^{-2x-6}}{5e^{-10}} = 2e^{-2x+4}$ for $2 < x$,

where $f(y) = 5e^{-5y}$ for $0 < y$

5-17. $c\int_0^\infty\int_x^\infty e^{-2x}e^{-3y}\,dy\,dx = \frac{c}{3}\int_0^\infty e^{-2x}(e^{-3x})\,dx = \frac{c}{3}\int_0^\infty e^{-5x}\,dx = \frac{1}{15}c$ $c = 15$

a)

$$P(X < 1, Y < 2) = 15\int_0^1\int_x^2 e^{-2x-3y}\,dy\,dx = 5\int_0^1 e^{-2x}(e^{-3x} - e^{-6})\,dx$$

$$= 5\int_0^1 e^{-5x}\,dx - 5e^{-6}\int_0^1 e^{-2x}\,dx = 1 - e^{-5} + \frac{5}{2}e^{-6}(e^{-2} - 1) = 0.9879$$

b) P(1 < X < 2) = $15\int_1^2\int_x^\infty e^{-2x-3y}dydx = 5\int_1^2 e^{-5x}dydx = (e^{-5} - e^{10}) = 0.0067$

c)

$$P(Y>3) = 15\left(\int_0^3\int_3^\infty e^{-2x-3y}dydx + \int_3^\infty\int_x^\infty e^{-2x-3y}dydx\right) = 5\int_0^3 e^{-9}e^{-2x}dx + 5\int_3^\infty e^{-5x}dx$$

$$= -\frac{3}{2}e^{-15} + \frac{5}{2}e^{-9} = 0.000308$$

d)

$$P(X<2, Y<2) = 15\int_0^2\int_x^2 e^{-2x-3y}dydx = 5\int_0^2 e^{-2x}(e^{-3x} - e^{-6})dx =$$

$$= 5\int_0^2 e^{-5x}dx - 5e^{-6}\int_0^2 e^{-2x}dx = \left(1-e^{-10}\right) + \frac{5}{2}e^{-6}\left(e^{-4}-1\right) = 0.9939$$

e) E(X) = $15\int_0^\infty\int_x^\infty xe^{-2x-3y}dydx = 5\int_0^\infty xe^{-5x}dx = \frac{1}{5^2} = 0.04$

f)

$$E(Y) = 15\int_0^\infty\int_x^\infty ye^{-2x-3y}dydx = \frac{-3}{2}\int_0^\infty 5ye^{-5y}dy + \frac{5}{2}\int_0^\infty 3ye^{-3y}dy$$

$$= -\frac{3}{10} + \frac{5}{6} = \frac{8}{15}$$

g) $f(x) = 15\int_x^\infty e^{-2x-3y}dy = \frac{15}{3}(e^{-2z-3x}) = 5e^{-5x}$ for $x > 0$

h) $f_X(1) = 5e^{-5}$ $\quad f_{XY}(1, y) = 15e^{-2-3y}$

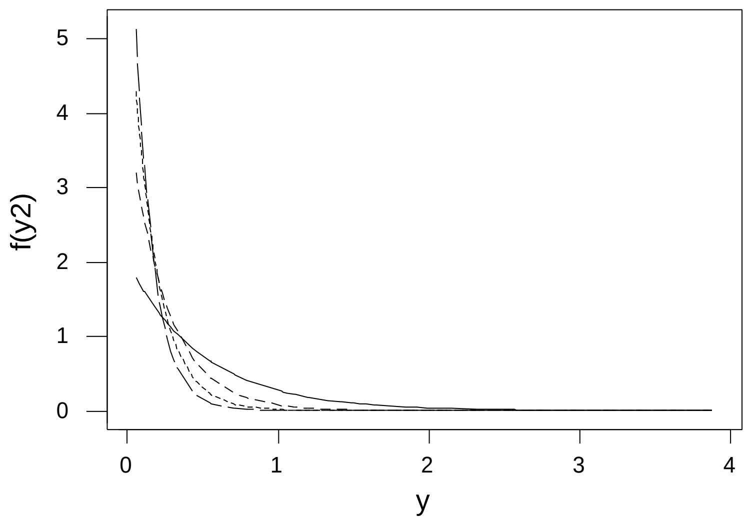
$f_{Y|X=1}(y) = \frac{15e^{-2-3y}}{5e^{-5}} = 3e^{3-3y}$ for 1 <y

i) $E(Y \mid X=1) = \int_1^\infty 3ye^{3-3y}dy = -ye^{3-3y}\Big|_1^\infty + \int_1^\infty e^{3-3y}dy = 4/3$

j) $\int_1^2 3e^{3-3y}dy = 1 - e^{-3} = 0.9502$ for $0 < y$, $f_Y(2) = \frac{15}{2}e^{-6}$

k) For $y > 0$ $f_{X|Y=2}(y) = \frac{15e^{-2x-6}}{\frac{15}{2}e^{-6}} = 2e^{-2x}$ for $0 < x < 2$

5-18. a) $f_{Y|X=x}(y)$, for x = 2, 4, 6, 8



b) $P(Y < 2 \mid X = 2) = \int_0^2 2e^{-2y} dy = 0.9817$

c) $E(Y \mid X = 2) = \int_0^\infty 2ye^{-2y} dy = 1/2$ (using integration by parts)

d) $E(Y \mid X = x) = \int_0^\infty xye^{-xy} dy = 1/x$ (using integration by parts)

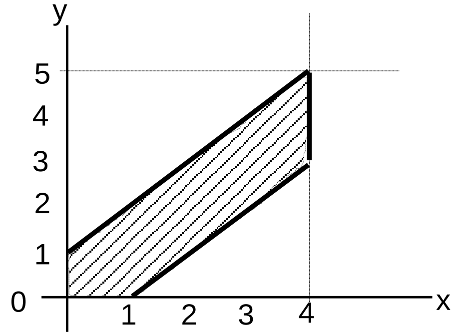
e) Use $f_X(x) = \dfrac{1}{b-a} = \dfrac{1}{10}$, $f_{Y|X}(x,y) = xe^{-xy}$, and the relationship

$$f_{Y|X}(x,y) = \frac{f_{XY}(x,y)}{f_X(x)}$$

Therefore, $xe^{-xy} = \dfrac{f_{XY}(x,y)}{1/10}$ and $f_{XY}(x,y) = \dfrac{xe^{-xy}}{10}$

f) $f_Y(y) = \int_0^{10} \dfrac{xe^{-xy}}{10} dx = \dfrac{1 - 10ye^{-10y} - e^{-10y}}{10y^2}$ (using integration by parts)

5-19. The graph of the range of (X, Y) is



$$\int_0^1 \int_0^{x+1} c\,dy\,dx + \int_1^4 \int_{x-1}^{x+1} c\,dy\,dx = 1$$

$$= c\int_0^1 (x+1)dx + 2c\int_1^4 dx$$

$$= \tfrac{3}{2}c + 6c = 7.5c = 1$$

Therefore, c = 1/7.5 = 2/15

a) $P(X < 0.5, Y < 0.5) = \int_0^{0.5}\int_0^{0.5} \tfrac{1}{7.5}\,dy\,dx = \tfrac{1}{30}$

b) $P(X < 0.5) = \int_0^{0.5}\int_0^{x+1} \tfrac{1}{7.5}\,dy\,dx = \tfrac{1}{7.5}\int_0^{0.5}(x+1)dx = \tfrac{2}{15}(\tfrac{5}{8}) = \tfrac{1}{12}$

c)

$$E(X) = \int_0^1\int_0^{x+1} \tfrac{x}{7.5}\,dy\,dx + \int_1^4\int_{x-1}^{x+1} \tfrac{x}{7.5}\,dy\,dx$$

$$= \tfrac{1}{7.5}\int_0^1 (x^2+x)dx + \tfrac{2}{7.5}\int_1^4 (x)dx = \tfrac{12}{15}(\tfrac{5}{6}) + \tfrac{2}{7.5}(7.5) = \tfrac{19}{9}$$

d)

$$E(Y) = \tfrac{1}{7.5}\int_0^1\int_0^{x+1} y\,dy\,dx + \tfrac{1}{7.5}\int_1^4\int_{x-1}^{x+1} y\,dy\,dx$$

$$= \tfrac{1}{7.5}\int_0^1 \tfrac{(x+1)^2}{2}dx + \tfrac{1}{7.5}\int_1^4 \tfrac{(x+1)^2-(x-1)^2}{2}dx$$

$$= \tfrac{1}{15}\int_0^1 (x^2+2x+1)dx + \tfrac{1}{15}\int_1^4 4x\,dx$$

$$= \tfrac{1}{15}\left(\tfrac{7}{3}\right) + \tfrac{1}{15}(30) = \tfrac{97}{45}$$

e)

$$f(x) = \int_0^{x+1} \frac{1}{7.5}dy = \left(\frac{x+1}{7.5}\right) \quad \text{for} \quad 0 < x < 1,$$

$$f(x) = \int_{x-1}^{x+1} \frac{1}{7.5}dy = \left(\frac{x+1-(x-1)}{7.5}\right) = \frac{2}{7.5} \quad \text{for} \quad 1 < x < 4$$

f)

$$f_{Y|X=1}(y) = \frac{f_{XY}(1,y)}{f_X(1)} = \frac{1/7.5}{2/7.5} = 0.5$$

$$f_{Y|X=1}(y) = 0.5 \quad \text{for } 0 < y < 2$$

g) $E(Y \mid X = 1) = \int_0^2 \frac{y}{2}dy = \left.\frac{y^2}{4}\right|_0^2 = 1$

h) $P(Y < 0.5 \mid X = 1) = \int_0^{0.5} 0.5dy = 0.5y\Big|_0^{0.5} = 0.25$

5-20. Let X, Y, and Z denote the time until a problem on line 1, 2, and 3, respectively.

a)

$$P(X > 40, Y > 40, Z > 40) = [P(X > 40)]^3$$

because the random variables are independent with the same distribution. Now,

$P(X > 40) = \int_{40}^{\infty} \frac{1}{40} e^{-x/40} dx = -e^{-x/40}\Big|_{40}^{\infty} = e^{-1}$ and the answer is

$$\left(e^{-1}\right)^3 = e^{-3} = 0.0498$$

b) $P(20 < X < 40, 20 < Y < 40, 20 < Z < 40) = [P(20 < X < 40)]^3$ and

$$P(20 < X < 40) = -e^{-x/40}\Big|_{20}^{40} = e^{-0.5} - e^{-1} = 0.2387 .$$

The answer is $0.2387^3 = 0.0136$

c) The joint density is not needed because the process is represented by three independent exponential distributions. Therefore, the probabilities may be multiplied.

5-21. $\mu = 3.2$, $\lambda = 1/3.2$

$$P(X > 5, Y > 5) = (1/10.24)\int_5^{\infty}\int_5^{\infty} e^{-\frac{x}{3.2}-\frac{y}{3.2}} dydx = 3.2\int_5^{\infty} e^{-\frac{x}{3.2}}\left(e^{-\frac{5}{3.2}}\right)dx$$

$$= \left(e^{-\frac{5}{3.2}}\right)\left(e^{-\frac{5}{3.2}}\right) = 0.0439$$

$$P(X > 10, Y > 10) = (1/10.24)\int_{10}^{\infty}\int_{10}^{\infty} e^{-\frac{x}{3.2}-\frac{y}{3.2}} dydx = 3.2\int_{10}^{\infty} e^{-\frac{x}{3.2}}\left(e^{-\frac{10}{3.2}}\right)dx$$

$$= \left(e^{-\frac{10}{3.2}}\right)\left(e^{-\frac{10}{3.2}}\right) = 0.0019$$

b) Let *X* denote the number of orders in a 5-minute interval. Then *X* is a Poisson random variable with $\lambda = 5/3.2 = 1.5625$.

$$P(X = 2) = \frac{e^{-1.5625}(1.5625)^2}{2!} = 0.256$$

For both systems, $P(X = 2)P(Y = 2) = 0.256^2 = 0.0655$

c) The joint probability distribution is not necessary because the two processes are independent and we can just multiply the probabilities.

5-22. (a) X: the life time of blade and Y: the life time of bearing

$f(x) = (1/3)e^{-x/3}$ $f(y) = (1/4)e^{-y/4}$

$P(X \geq 5, Y \geq 5) = P(X \geq 5)P(Y \geq 5) = e^{-5/3}e^{-5/4} = 0.0541$

(b) P(X > t, Y > t) = $e^{-t/3}e^{-t/4} = e^{-7t/12} = 0.95 \rightarrow t = -12\ln(0.95)/7 = 0.0879$ years

5-23. a) $P(X<0.5)=\int_0^{0.5}\int_0^1\int_0^1(8xyz)dzdydx=\int_0^{0.5}\int_0^1(4xy)dydx=\int_0^{0.5}(2x)dx=x^2\Big|_0^{0.5}=0.25$

b)

$$P(X<0.5,Y<0.5)=\int_0^{0.5}\int_0^{0.5}\int_0^1(8xyz)dzdydx$$

$$=\int_0^{0.5}\int_0^{0.5}(4xy)dydx=\int_0^{0.5}(0.5x)dx=\tfrac{x^2}{4}\Big|_0^{0.5}=0.0625$$

c) P(Z < 2) = 1, because the range of Z is from 0 to 1

d) P(X < 0.5 or Z < 2) = P(X < 0.5) + P(Z < 2) - P(X < 0.5, Z < 2).
Now, P(Z < 2) = 1 and P(X < 0.5, Z < 2) = P(X < 0.5). Therefore, the answer is 1.

e) $E(X)=\int_0^1\int_0^1\int_0^1(8x^2yz)dzdydx=\int_0^1(2x^2)dx=\tfrac{2x^3}{3}\Big|_0^1=2/3$

f) $P(X<0.5|Y=0.5)$ is the integral of the conditional density $f_{X|Y}(x)$. Now,

$$f_{X|0.5}(x)=\frac{f_{XY}(x,0.5)}{f_Y(0.5)} \quad \text{and} \quad f_{XY}(x,0.5)=\int_0^1(8x(0.05)z)dz=4x0.5=2x$$

for 0 < x < 1 and 0 < y < 1.

Also, $f_Y(y)=\int_0^1\int_0^1(8xyz)dzdx=2y$ for 0 < y < 1.

Therefore, $f_{X|0.5}(x)=\frac{2x}{1}=2x$ for 0 < x < 1. Then, $P(X<0.5|Y=0.5)=\int_0^{0.5}2xdx=0.25$

.

g) $P(X<0.5,Y<0.5|Z=0.8)$ is the integral of the conditional density of X and Y.

Now, $f_Z(z)=2z$ for 0 < z < 1 as in part a) and

$$f_{XY|Z}(x,y)=\frac{f_{XYZ}(x,y,z)}{f_Z(z)}=\frac{8xy(0.8)}{2(0.8)}=4xy \text{ for } 0<x<1 \text{ and } 0<y<1.$$

Then, $P(X<0.5,Y<0.5|Z=0.8)=\int_0^{0.5}\int_0^{0.5}(4xy)dydx=\int_0^{0.5}(x/2)dx=\tfrac{1}{16}=0.0625$

h) $f_{YZ}(y,z)=\int_0^1(8xyz)dx=4yz$ for 0 < y < 1 and 0 < z < 1.

Then, $f_{X|YZ}(x)=\frac{f_{XYZ}(x,y,z)}{f_{YZ}(y,z)}=\frac{8x(0.5)(0.8)}{4(0.5)(0.8)}=2x$ for 0 < x < 1.

i) Therefore, $P(X<0.5|Y=0.5,Z=0.8)=\int_0^{0.5}2xdx=0.25$

5-24.

$\iint_{x^2+y^2\le 4}\int_0^4 c\,dz\,dy\,dx$ = the volume of a cylinder with a base of radius 2 and a height of 4 =

$(\pi 2^2)4 = 16\pi$. Therefore, $c = \dfrac{1}{16\pi}$

a) $P(X^2 + Y^2 < 2)$ equals the volume of a cylinder of radius $\sqrt{2}$ and a height of 4 (= 8π) times *c*. Therefore, the answer is $\dfrac{8\pi}{16\pi} = 1/2.$

b) P(Z < 2) equals half the volume of the region where $f_{XYZ}(x, y, z)$ is positive times 1/c. Therefore, the answer is 0.5.

c) $E(X) = \int_{-2}^{2}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}}\int_0^4 \frac{x}{c}\,dz\,dy\,dx = c\int_{-2}^{2}\left[4xy\Big|_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}}\right]dx = c\int_{-2}^{2}(8x\sqrt{4-x^2})dx$.

Using substitution, $u = 4 - x^2$, $du = -2x\,dx$, and

$$E(X) = c\int 4\sqrt{u}\,du = \tfrac{-4}{c}\tfrac{2}{3}(4-x^2)^{\frac{3}{2}}\Big|_{-2}^{2} = 0$$

d) $f_{X|1}(x) = \dfrac{f_{XY}(x,1)}{f_Y(1)}$ *and* $f_{XY}(x, y) = c\int_0^4 dz = \frac{4}{c} = \frac{1}{4\pi}$ *for* $x^2 + y^2 < 4$.

Also, $f_Y(y) = c\int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}}\int_0^4 dz\,dx = 8c\sqrt{4-y^2}$ for -2 < y < 2.

Then, $f_{X|y}(x) = \dfrac{4c}{8c\sqrt{4-y^2}}$ evaluated at y = 1. That is, $f_{X|1}(x) = \frac{1}{2\sqrt{3}}$ for $-\sqrt{3} < x < \sqrt{3}$

Therefore, $P(X < 1 \mid Y < 1) = \int_{-\sqrt{3}}^{1} \frac{1}{2\sqrt{3}}\,dx = \dfrac{1+\sqrt{3}}{2\sqrt{3}} = 0.7887$

e) $f_{XY|1}(x, y) = \dfrac{f_{XYZ}(x, y, 1)}{f_Z(1)}$ *and* $f_Z(z) = \int_{-2}^{2}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} c\,dy\,dx = \int_{-2}^{2} 2c\sqrt{4-x^2}\,dx$

Because $f_Z(z)$ is a density over the range 0 < z < 4 that does not depend on Z, $f_Z(z)$ =1/4 for 0 < z < 4. Then, $f_{XY|1}(x, y) = \dfrac{c}{1/4} = \dfrac{1}{4\pi}$ for $x^2 + y^2 < 4$.

Then, $P(X^2 + Y^2 < 1 \mid Z = 1) = \dfrac{\text{area in } x^2 + y^2 < 1}{4\pi} = 1/4$

f) $f_{Z|xy}(z) = \dfrac{f_{XYZ}(x, y, z)}{f_{XY}(x, y)}$ and $f_{XY}(x, y) = \frac{1}{4\pi}$ *for* $x^2 + y^2 < 4$. Therefore,

$$f_{Z|xy}(z) = \frac{\frac{1}{16\pi}}{\frac{1}{4\pi}} = 1/4 \text{ for } 0 < z < 4.$$

5-25. Determine c such that $f(xyz) = c$ is a joint density probability over the region x > 0, y > 0 and z > 0 with x + y + z < 1

$$f(xyz) = c\int_0^1\int_0^{1-x}\int_0^{1-x-y} dzdydx = \int_0^1\int_0^{1-x} c(1-x-y)dydx = \int_0^1\left(c(y-xy-\frac{y^2}{2})\Big|_0^{1-x}\right)dx$$

$$= \int_0^1 c\left((1-x)-x(1-x)-\frac{(1-x)^2}{2}\right)dx = \int_0^1 c\left(\frac{(1-x)^2}{2}\right)dx = c\left(\frac{1}{2}x-\frac{x^2}{2}+\frac{x^3}{6}\right)\Big|_0^1$$

$$= c\frac{1}{6}. \quad \text{Therefore, } c = 6.$$

a) $P(X<0.5, Y<0.5, Z<0.5) = 6\int_0^1\int_0^{1-x}\int_0^{1-x-y} dzdydx \Rightarrow$ The conditions $x<0.5$, $y<0.5$, $z<0.5$ and $x+y+z<1$ make a space that is a cube with a volume of 0.125. Therefore the probability of $P(X<0.5, Y<0.5, Z<0.5) = 6(0.125) = 0.75$

b)

$$P(X<0.5, Y<0.5) = \int_0^{0.5}\int_0^{0.5} 6(1-x-y)dydx = \int_0^{0.5}\left(6y-6xy-3y^2\right)\Big|_0^{0.5} dx$$

$$= \int_0^{0.5}\left(\frac{9}{4}-3x\right)dx = \left(\frac{9}{4}x-\frac{3}{2}x^2\right)\Big|_0^{0.5} = 3/4$$

c)

$$P(X<0.5) = 6\int_0^{0.5}\int_0^{1-x}\int_0^{1-x-y} dzdydx = \int_0^{0.5}\int_0^{1-x} 6(1-x-y)dydx = \int_0^{0.5} 6(y-xy-\frac{y^2}{2})\Big|_0^{1-x}$$

$$= \int_0^{0.5} 6(\frac{x^2}{2}-x+\frac{1}{2})dx = \left(x^3-3x^2+3x\right)\Big|_0^{0.5} = 0.875$$

d)

$$E(X) = 6\int_0^1\int_0^{1-x}\int_0^{1-x-y} xdzdydx = \int_0^1\int_0^{1-x} 6x(1-x-y)dydx = \int_0^1 6x(y-xy-\frac{y^2}{2})\Big|_0^{1-x}$$

$$= \int_0^1 6(\frac{x^3}{2}-x^2+\frac{x}{2})dx = \left(\frac{3x^4}{4}-2x^3+\frac{3x^2}{2}\right)\Big|_0^1 = 0.25$$

e)

$$f(x) = 6\int_0^{1-x}\int_0^{1-x-y} dzdy = \int_0^{1-x} 6(1-x-y)dy = 6\left(y-xy-\frac{y^2}{2}\right)\Big|_0^{1-x}$$

$$= 6(\frac{x^2}{2}-x+\frac{1}{2}) = 3(x-1)^2 \text{ for } 0<x<1$$

f)

$$f(x,y) = 6\int_0^{1-x-y} dz = 6(1-x-y)$$

for $x > 0,\ y > 0\ and\ x + y < 1$

g)

$$f(x \mid y = 0.5, z = 0.5) = \frac{f(x, y = 0.5, z = 0,5)}{f(y = 0.5, z = 0.5)} = \frac{6}{6} = 1 \text{ for } x > 0$$

h) The marginal $f_Y(y)$ is similar to $f_X(x)$ and $f_Y(y) = 3(1-y)^2$ for $0 < y < 1$.

$$f_{X|Y}(x \mid 0.5) = \frac{f(x,0.5)}{f_Y(0.5)} = \frac{6(0.5-x)}{3(0.25)} = 4(1-2x) \text{ for } x < 0.5$$

5-26. Let X denote the production yield on a day. Then,

$$P(X > 1400) = P(Z > \tfrac{1400-1500}{\sqrt{10000}}) = P(Z > -1) = 0.84134.$$

a) Let Y denote the number of days out of five such that the yield exceeds 1400. Then, by independence, Y has a binomial distribution with $n = 5$ and $p = 0.8413$. Therefore, the answer is

$$P(Y = 5) = \binom{5}{5} 0.8413^5 (1-0.8413)^0 = 0.4215$$

b) As in part (a), the answer is

$$P(Y \ge 4) = P(Y = 4) + P(Y = 5) = \binom{5}{4} 0.8413^4 (1-0.8413)^1 + 0.4215 = 0.8190$$

5-27.

a) Let X denote the weight of a brick. Then,

$$P(X > 2.75) = P(Z > \tfrac{2.75-3}{0.25}) = P(Z > -1) = 0.84134$$

Let Y denote the number of bricks in the sample of 20 that exceed 2.75 pounds. Then, by independence, Y has a binomial distribution with n = 20 and p = 0.84134. Therefore, the answer is

$$P(Y = 20) = \binom{20}{20} 0.84134^{20} = 0.032.$$

b) Let A denote the event that the heaviest brick in the sample exceeds 3.75 pounds. Then, P(A) = 1 - P(A') and A' is the event that all bricks weigh less than 3.75 pounds. As in part a),

P(X < 3.75) = P(Z < 3) and $P(A) = 1 - [P(Z < 3)]^{20} = 1 - 0.99865^{20} = 0.0267$

5-28. a) Let X denote the grams of luminescent ink. Then,

$$P(X < 1.14) = P(Z < \tfrac{1.14-1.2}{0.3}) = P(Z < -2) = 0.022750$$

Let Y denote the number of bulbs in the sample of 25 that have less than 1.14 grams. Then, by independence, Y has a binomial distribution with n = 25 and p = 0.022750. Therefore, the answer is $P(Y \ge 1) = 1 - P(Y = 0) = \binom{25}{0} 0.02275^0 (0.97725)^{25} = 1 - 0.5625 = 0.4375$.

b)

$$\begin{aligned}
P(Y \le 5) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + (P(Y = 3) + P(Y = 4) + P(Y = 5) \\
&= \binom{25}{0} 0.02275^0 (0.97725)^{25} + \binom{25}{1} 0.02275^1 (0.97725)^{24} + \binom{25}{2} 0.02275^2 (0.97725)^{23} \\
&\quad + \binom{25}{3} 0.02275^3 (0.97725)^{22} + \binom{25}{4} 0.02275^4 (0.97725)^{21} + \binom{25}{5} 0.02275^5 (0.97725)^{20} \\
&= 0.5625 + 0.3274 + 0.09146 + 0.01632 + 0.002090 + 0.0002043 = 0.99997 \cong 1
\end{aligned}$$

c) $P(Y=0)=\binom{25}{0}0.02275^0(0.97725)^{25}=0.5625$

d) The lamps are normally and independently distributed. Therefore, the probabilities can be multiplied.

5-29. a)

$f_{Y|x}(y)=e^{-(y-x)}\ge 0$ for all $y > x$.

$$\int_x^\infty f_{Y|x}(y)dy=\int_x^\infty e^{-(y-x)}dy=e^x\int_x^\infty e^{-y}dy=e^x(e^{-x})=1$$

$P(Y\in B\mid X=x)=\int_B f_{Y|x}(y)dy$ for any set B in the range of Y.

As a result, $f_{Y|x}(y)$ is a probability density function for any value of x.

b) A joint probability density function only has nonzero probability where the conditional probability is nonzero, namely the region $x < y$. Therefore, $P(X<Y)=1$ for $0<x<y$

c) $f_{XY}(x,y)=f_X(x)f_{Y|x}(y)=e^{-x}\times e^{-(y-x)}=e^{-y}$ for $0<x<y$

d) $f_{X|y}(x)=\dfrac{f_{XY}(x,y)}{f_Y(y)}$ where $f_Y(y)=\int_0^y e^{-y}dx=ye^{-y}$ and this implies

$$f_{X|y}(x)=\frac{e^{-y}}{ye^{-y}}=\frac{1}{y} \text{ for } 0<x<y$$

e)

$P(Y<2\mid X=1)=\int_1^2 f_{Y|1}(y)dy$ because $x<y$ and $x=1$. Therefore

$$P(Y<2\mid X=1)=\int_1^2 e^{-(y-1)}dy=1-e^{-1}$$

f)

$$E(Y\mid X=1)=\int_1^\infty ye^{-(y-1)}dy=e\int_1^\infty ye^{-y}dy=e\left[-ye^{-y}\Big|_1^\infty-\int_1^\infty -e^{-y}dy\right]=e\left[\frac{1}{e}+\frac{1}{e}\right]=2 \text{ from}$$

integration by parts

g)

$$P(X<1,Y<1)=\int_0^1\int_x^1 f_{XY}(x,y)dydx=\int_0^1\int_x^1 e^{-y}dydx=\int_0^1(-e^{-1}+e^{-x})dx=1-\frac{2}{e}$$

h)

$$P(Y<2)=\int_0^2 ye^{-y}dy=-ye^{-y}\Big|_0^2-\int_0^2 -e^{-y}dy=1-3e^{-2} \text{ from integration by parts}$$

i)

Solve for c with solver software

$$P(Y<c)=\int_0^c ye^{-y}dy=-ye^{-y}\Big|_0^c-\int_0^c -e^{-y}dy=1-(c+1)e^{-c}=0.9$$

c = 3.9

j) X and Y are not independent. For example, $f_{Y|x}(y) \neq f_Y(y)$.

5-30.

	Y		
X	0	50	75
0	0.9819	0.0122	0.0059
50	0.1766	0.7517	0.0717
75	0.0237	0.0933	0.883

$P(X = 75) = 0.9$, $P(X= 50) = 0.08$, $P(X=0) = 0.02$

a)

$$P(Y \le 50 \mid X = 50) = P(Y = 0 \mid X = 50) + P(Y = 50 \mid X = 50)$$
$$= 0.1766 + 0.7517 = 0.9283$$

b)

$$P(X = 0, Y = 75) = P(X = 0)P(Y = 75 \mid X = 0)$$
$$= 0.02 \times 0.0059 = 0.000118$$

c)

$$E(Y \mid X = 50) = \sum_y y \times P(Y = y \mid X = 50)$$
$$= 0 + 50 \times 0.7517 + 75 \times 0.0717$$
$$= 42.9625$$

d) $f_Y(y) = \sum_x P(Y = y \mid X = x)P(X = x)$

$$f_Y(y) = \begin{cases} 0.9819 \times 0.02 + 0.1766 \times 0.08 + 0.0237 \times 0.9 = 0.0551, & y = 0 \\ 0.0122 \times 0.02 + 0.7517 \times 0.08 + 0.0933 \times 0.9 = 0.1444, & y = 50 \\ 0.0059 \times 0.02 + 0.0717 \times 0.08 + 0.8830 \times 0.9 = 0.8006, & y = 75 \end{cases}$$

e)

$$f_{XY}(x, y) = P(Y = y \mid X = x)P(X = x)$$

$$f_{XY}(x, y) = \begin{cases} 0.9819 \times 0.02 = 0.019638, & x = 0, y = 0 \\ 0.0122 \times 0.02 = 0.000244, & x = 0, y = 50 \\ 0.0059 \times 0.02 = 0.000118, & x = 0, y = 75 \\ 0.1766 \times 0.08 = 0.014128, & x = 50, y = 0 \\ 0.7517 \times 0.08 = 0.060136, & x = 50, y = 50 \\ 0.0717 \times 0.08 = 0.005736 & x = 50, y = 75 \\ 0.0237 \times 0.9 = 0.02133, & x = 75, y = 0 \\ 0.0933 \times 0.9 = 0.08397, & x = 75, y = 50 \\ 0.8830 \times 0.9 = 0.7947, & x = 75, y = 75 \end{cases}$$

f) X and Y are not independent because knowledge of the values of X changes the probabilities associated with the values for Y. For example, $P(Y = 0 \mid X = 0) \neq P(Y = 0 \mid X = 50)$.

5-31.

X: Demand for MMR vaccine is normally distributed with mean 1.1 and standard deviation 0.3.
Y: Demand for varicella vaccine is normally distributed with mean 0.55 and standard deviation 0.1.

a) $P(X \le 1.2, Y \le 0.6) = P(X \le 1.2)P(Y \le 0.6)$ because *X* and *Y* are independent.

$$= P\left(\frac{X-1.1}{0.3} \le \frac{1.2-1.1}{0.3}\right)P\left(\frac{Y-0.55}{0.1} \le \frac{0.6-0.55}{0.1}\right) = (0.6293)(0.6915) = 0.4352$$

b)

$$P(X \le x, Y \le y) = 0.90$$

$$\Rightarrow P\left(\frac{X-1.1}{0.3} \le \frac{x-1.1}{0.3}\right)P\left(\frac{Y-0.55}{0.1} \le \frac{y-0.55}{0.1}\right) = 0.9$$

$$\Rightarrow P\left(Z_1 \le \frac{x-1.1}{0.3}\right)P\left(Z_2 \le \frac{y-0.55}{0.1}\right) = 0.9$$

where *x*, and *y* are the inventory levels for MMR, and varicella vaccines, respectively. Any combination of *x*, and *y* that satisfies the above condition is a solution. A possible solution is, $x = 1.625$ and $y = 0.703$

5-32.

a) $f_{Y|73}(y) = \frac{1}{10\sqrt{2\pi}} e^{\frac{-(y-(1.6\times 73))^2}{2(10^2)}}$ for $X = 73$

b) $P(Y < 115 \mid X = 73) = P\left(Z \le \frac{115-(1.6\times 73)}{10}\right) = P(Z \le -0.18) = 0.4286$

c) $E(Y \mid X = 73) = 1.6 \times 73 = 116.8$

d)

$$f_{X,Y}(x.y) = f_{Y|x}(y) f_X(x) = \frac{1}{10\sqrt{2\pi}} e^{\frac{-(y-1.6x)^2}{2(10^2)}} \frac{1}{8\sqrt{2\pi}} e^{\frac{-(y-73)^2}{2(8^2)}}$$

$$= \frac{1}{160\pi} e^{\frac{-(y-1.6x)^2}{2(10^2)} + \frac{-(y-73)^2}{2(8^2)}}$$

This is recognized as a bivariate normal distribution. From the formulas for the mean and variance of a conditional normal distribution we have

$$(1-\rho^2)\sigma_Y^2 = 100$$

$$\mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(x-\mu_X) = 1.6x$$

From the second equation

$$\frac{\sigma_Y}{\sigma_X}\rho = 1.6 \text{ and } \mu_Y - \frac{\sigma_Y}{\sigma_X}\rho\mu_X = 0$$

and these can be written as

$$\sigma_Y = \frac{1.6\sigma_X}{\rho} \text{ and } \mu_Y = \frac{\sigma_Y}{\sigma_X}\rho\mu_X$$

From the first equation above

$$(1-\rho^2)\left(\frac{1.6^2\sigma_X^2}{\rho^2}\right) = 100$$

Because $\sigma_X = 8$, the previous equation can be solved for ρ to yield

$$(1-\rho^2)\left(\frac{163.84}{\rho^2}\right) = 100$$

and ρ = 0.788

Then $\sigma_Y = \frac{1.6\sigma_X}{\rho} = \frac{12.8}{0.788} = 16.24$ and $\sigma_Y^2 = 263.84$

Also, $\mu_Y = \frac{\sigma_Y}{\sigma_X}\rho\mu_X = \frac{16.24}{8}(0.788)(73) = 116.8$

Section 5-2

5-33. E(X) = 1(3/8) + 2(1/2) + 4(1/8) = 15/8 = 1.875
E(Y) = 3(1/8) + 4(1/4) + 5(1/2) + 6(1/8) = 37/8 = 4.625

$$\begin{aligned}\mathrm{E(XY)} &= [1\times 3\times(1/8)] + [1\times 4\times(1/4)] + [2\times 5\times(1/2)] + [4\times 6\times(1/8)] \\ &= 75/8 = 9.375\end{aligned}$$

$$\sigma_{XY} = E(XY) - E(X)E(Y) = 9.375 - (1.875)(4.625) = 0.703125$$

$V(X) = 1^2(3/8) + 2^2(1/2) + 4^2(1/8) - (15/8)^2 = 0.8594$

$V(Y) = 3^2(1/8) + 4^2(1/4) + 5^2(1/2) + 6^2(1/8) - (37/8)^2 = 0.7344$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{0.703125}{\sqrt{(0.8594)(0.7344)}} = 0.8851$$

5-34.
$$E(X) = -1(1/8) + (-0.5)(1/4) + 0.5(1/2) + 1(1/8) = 0.125$$
$$E(Y) = -2(1/8) + (-1)(1/4) + 1(1/2) + 2(1/8) = 0.25$$
$$\mathrm{E(XY)} = [-1\times -2\times(1/8)] + [-0.5\times -1\times(1/4)] + [0.5\times 1\times(1/2)] + [1\times 2\times(1/8)] = 0.875$$

V(X) = 0.4219
V(Y) = 1.6875

$$\sigma_{XY} = 0.875 - (0.125)(0.25) = 0.8438$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{0.8438}{\sqrt{0.4219}\sqrt{1.6875}} = 1$$

5-35.

$$\sum_{x=1}^{3}\sum_{y=1}^{3} c(x+y) = 36c, \qquad c = 1/36$$

$$E(X) = \frac{13}{6} \qquad E(Y) = \frac{13}{6} \qquad E(XY) = \frac{14}{3} \qquad \sigma_{xy} = \frac{14}{3} - \left(\frac{13}{6}\right)^2 = \frac{-1}{36}$$

$$E(X^2) = \frac{16}{3} \qquad E(Y^2) = \frac{16}{3} \qquad V(X) = V(Y) = \frac{23}{36}$$

$$\rho = \frac{\frac{-1}{36}}{\sqrt{\frac{23}{36}}\sqrt{\frac{23}{36}}} = -0.0435$$

5-36.

The marginal distribution of X is

x	f(x)
0	0.75
1	0.2
2	0.05

$$E(X)=0(0.75)+1(0.2)+2(0.05)=0.3$$
$$E(Y)=0(0.3)+1(0.28)+2(0.25)+3(0.17)=1.29$$
$$E(X^2)=0(0.75)+1(0.2)+4(0.05)=0.4$$
$$E(Y^2)=0(0.3)+1(0.28)+4(0.25)+9(0.17)=1.146$$

V(X) = $0.4 - 0.3^2 = 0.31$
V(Y) = $2.81 - 1.146^2 = 1.16$

$$\mathrm{E(XY)}=[0\times 0\times(0.225)]+[0\times 1\times(0.21)]+[0\times 2\times(0.1875)]+...+[2\times 3\times(0.0085)]=0.387$$

$$\sigma_{XY}=0.387-(0.3)(1.29)=0$$
$$\rho_{XY}=\tfrac{\sigma_{XY}}{\sigma_X\sigma_Y}=0$$

5-37.

Let *X* and *Y* denote the number of patients who improve or degrade, respectively, and let *Z* denote the number of patients that remain the same. If *X* = 0, then *Y* can equal 0,1,2,3, or 4. However, if *X* = 4 then *Y* = 0. Consequently, the range of the joint distribution of *X* and *Y* is not rectangular. Therefore, *X* and *Y* are not independent.

Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y).
Therefore,
Cov(X,Y) = 0.5[Var(X + Y) - Var(X) - Var(Y)]

Here *X* and *Y* are binomially distributed when considered individually. Therefore,

$$f_X(x)=\frac{4!}{x!(4-x)!}0.4^x(1-0.4)^{4-x}$$
$$f_Y(y)=\frac{4!}{y!(4-y)!}0.1^y(1-0.1)^{4-y}$$

And
Var(X) = 4(0.4)(0.6) = 0.96
Var(Y) = 4(0.1)(0.9) = 0.36

Also, $W = X + Y$ is binomial with $n = 4$, and $p = 0.4 + 0.1 = 0.5$. Therefore,
Var($X + Y$) = 4(0.5)(0.5) = 1

Therefore, Cov(X,Y) = 0.5[1 – 0.96 – 0.36] = –0.16

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}=\frac{-0.16}{\sqrt{0.96\times 0.36}}=-0.272$$

5-38.

Transaction	Frequency	Selects(X)	Updates(Y)	Inserts(Z)
New Order	43	23	11	12
Payment	44	4.2	3	1
Order Status	4	11.4	0	0
Delivery	5	130	120	0
Stock Level	4	0	0	0
Mean Value		18.694	12.05	5.6

a) Cov(X,Y) = E(XY) - E(X)E(Y) = 23(11)(0.43) + 4.2(3)(0.44) + 11.4(0)(0.04) + 130(120)(0.05) + 0(0)(0.04) - 18.694(12.05) = 669.0713

b) V(X) = 735.9644, V(Y) = 630.7875, Corr(X,Y) = cov(X,Y)/(V(X) V(Y))$^{0.5}$ = 0.9820

c) Cov(X,Z)=23(12)(0.43) + 4.2(1)(0.44) + 0 - 18.694(5.6) = 15.8416

d) V(Z)=31, Corr(X,Z) = 0.1049

5-39. Here c = 8/31

$$E(XY) = \frac{8}{81}\int_0^3\int_0^x xy(xy)dydx = \frac{8}{81}\int_0^3\int_0^x x^2y^2dydx = \frac{8}{81}\int_0^3 \frac{x^3}{3}x^2dx = \frac{8}{81}\int_0^3 \frac{x^5}{3}dx$$

$$= \left(\frac{8}{81}\right)\left(\frac{3^6}{18}\right) = 4$$

$$\sigma_{xy} = 4 - \left(\frac{12}{5}\right)\left(\frac{8}{5}\right) = 0.16$$

$$E(X^2) = 6 \qquad E(Y^2) = 3$$

$$V(x) = 0.24, \qquad V(Y) = 0.44$$

$$\rho = \frac{0.16}{\sqrt{0.24}\sqrt{0.44}} = 0.4924$$

5-40. Here $c = 2/19$

$$E(X) = \frac{2}{19}\int_0^1\int_0^{x+1} xdydx + \frac{2}{19}\int_1^5\int_{x-1}^{x+1} xdydx = 2.614$$

$$E(Y) = \frac{2}{19}\int_0^1\int_0^{x+1} ydydx + \frac{2}{19}\int_1^5\int_{x-1}^{x+1} ydydx = 2.649$$

Now, $E(XY) = \frac{2}{19}\int_0^1\int_0^{x+1} xydydx + \frac{2}{19}\int_1^5\int_{x-1}^{x+1} xydydx = 8.7763$

$$\sigma_{xy} = 8.7763 - (2.614)(2.649) = 1.85181$$
$$E(X^2) = 8.7632 \qquad E(Y^2) = 9.11403$$
$$V(x) = 1.930, \qquad V(Y) = 2.0968$$
$$\rho = \frac{1.852}{\sqrt{1.930}\sqrt{2.097}} = 0.9206$$

5-41. a) E(X) = 1 E(Y) = 1

$$E(XY) = \int_0^\infty\int_0^\infty xye^{-x-y}dxdy = \int_0^\infty xe^{-x}dx\int_0^\infty ye^{-y}dy = E(X)E(Y)$$

Therefore, $\sigma_{XY} = \rho_{XY} = 0$.

5-42.

E(X) = 333.33, E(Y) = 833.33
$E(X^2)$ = 222,222.2
V(X) = 222222.2-$(333.33)^2$ = 111,113.31
$E(Y^2)$ = 1,055,556
V(Y) = 361,117.11

$$E(XY) = 6\times10^{-6}\int_0^\infty\int_x^\infty xye^{-.001x-.002y}dydx = 388,888.9$$
$$Cov(X,Y) = 388,888.9 - (333.33)(833.33) = 111,115.01$$
$$\rho = \frac{111,115.01}{\sqrt{111113.31}\sqrt{361117.11}} = 0.5547$$

5-43. $E(X) = -1(1/4) + 1(1/4) = 0$

$$E(Y) = -1(1/4) + 1(1/4) = 0$$
$$\text{E(XY)} = [-1\times0\times(1/4)] + [-1\times0\times(1/4)] + [1\times0\times(1/4)] + [0\times1\times(1/4)] = 0$$

V(X) = 1/2
V(Y) = 1/2

$$\sigma_{XY} = 0 - (0)(0) = 0$$
$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{0}{\sqrt{1/2}\sqrt{1/2}} = 0$$

The correlation is zero, but *X* and *Y* are not independent. For example, if y = 0, then x must be –1 or 1.

5-44.

$$\mu_X = \sum_x xP(X=x) = 0 + 50\times0.08 + 75\times0.9 = 71.5$$

x	[x-E(X)]^2	P(X=x)	Product
0	5112.25	0.02	102.245
50	462.25	0.08	36.98
75	12.25	0.9	11.025
			V(X) =150.25

$$V(X) = E(X-\mu_X)^2 = \sum_x (x-\mu_X)^2 P(X=x) = 150.25$$

From the solution for the referenced exercise, we know the probability distribution for Y

$$f_Y(y)=\begin{cases}0.9819\times 0.02+0.1766\times 0.08+0.0237\times 0.9=0.0551, & y=0\\ 0.0122\times 0.02+0.7517\times 0.08+0.0933\times 0.9=0.1444, & y=50\\ 0.0059\times 0.02+0.0717\times 0.08+0.8830\times 0.9=0.8006, & y=75\end{cases}$$

$$\mu_Y=\sum_y yP(Y=y)=0+50\times 0.14435+75\times 0.8006=67.259$$

y	*E*(*Y*)	[*y* - *E*(*Y*)]²	*P*(*Y*=*y*)	*P*(*Y*=*y*)[*y* - *E*(*Y*)]²
0	67.25905	4523.7798	0.055096	249.2422
50	67.25905	297.87481	0.14435	42.99823
75	67.25905	59.922307	0.800554	47.97104
				340.2114

$$V(Y)=E(Y-\mu_Y)^2=\sum_y (y-\mu_Y)^2 P(Y=y)=340.211$$

$$f_{XY}(x,y)=\begin{cases}0.9819\times 0.02=0.019638, & x=0, y=0\\ 0.01222\times 0.02=0.000244, & x=0, y=50\\ 0.0059\times 0.02=0.000118, & x=0, y=75\\ 0.1766\times 0.08=0.014128, & x=50, y=0\\ 0.7517\times 0.08=0.060136, & x=50, y=50\\ 0.0717\times 0.08=0.005736 & x=50, y=75\\ 0.0237\times 0.9=0.02133, & x=75, y=0\\ 0.0933\times 0.9=0.08397, & x=75, y=50\\ 0.8830\times 0.9=0.7947, & x=75, y=75\end{cases}$$

x	y	[x-E(X)][y-E(Y)]	P(X=x, Y=y)	Product
0	0	4809.022075	0.019638	94.43958
0	50	1234.022075	0.000244	0.301101
0	75	-553.477925	0.000118	-0.06531
50	0	1446.069575	0.014128	20.43007
50	50	371.069575	0.060136	22.31464
50	75	-166.430425	0.005736	-0.95464
75	0	-235.406675	0.02133	-5.02122
75	50	-60.406675	0.08397	-5.07235
75	75	27.093325	0.7947	21.53107
				147.9029

$$Cov(X,Y)=E((X-\mu_X)(Y-\mu_Y))=\sum_x\sum_y (x-\mu_X)(y-\mu_Y)f_{XY}(x,y)$$

$$=(0-71.5)(0-67.25905)0.019638+\ldots$$

$$+(75-71.5)(75-67.25905)0.7947=147.903$$

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}=\frac{138.531418}{\sqrt{150.25\times 340.211}}=0.654$$

5-45. $\mu_X = \int_0^\infty xe^{-x}dx = -xe^{-x}\Big|_0^\infty + \int_0^\infty e^{-x}dx = \left(-x^{-x}\Big|_0^\infty\right) + \left(-e^{-x}\Big|_0^\infty\right) = 0 + 1 = 1$

The probability density function for *Y* was determined in the solution to the referenced exercise.

$$\mu_Y = \int_0^\infty y(ye^{-y})dy = \int_0^\infty y^2e^{-y})dy = -y^2e^{-y}\Big|_0^\infty - \int_0^\infty 2y(-e^{-y})dy$$

$$= -y^2e^{-y}\Big|_0^\infty + 2\left(-ye^{-y}\Big|_0^\infty - \int_0^\infty -e^{-y}dy\right) = -y^2e^{-y}\Big|_0^\infty + 2\left(-ye^{-y}\Big|_0^\infty - e^{-y}\Big|_0^\infty\right)$$

$$= 0 + 2(0+1) = 2$$

$$E(XY) = \int_0^\infty\int_0^y xye^{-y}dxdy = \int_0^\infty ye^{-y}\left(\frac{x^2}{2}\right)\Big|_0^y dy = \int_0^\infty ye^{-y}\frac{y^2}{2}dy = \frac{1}{2}\int_0^\infty y^3e^{-y}dy$$

Using integration by parts multiple times

…

$$= \frac{1}{2}\left(-y^3e^{-y} - 3y^2e^{-y} - 6ye^{-y} - 6e^{-y}\right)\Big|_0^\infty = \frac{1}{2}\left(0-(-6)\right) = 3$$

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y = 3 - 1\times 2 = 1$$

$$V(X) = \int_0^\infty x^2 f_X(x)dx - \mu_X^2 = \int_0^\infty x^2e^{-x}dx - \mu_X^2$$

$$= -x^2e^{-x}\Big|_0^\infty + \int_0^\infty 2xe^{-x}dx - \mu_X^2 = 0 + 2(1) - 1^2 = 1$$

$$V(Y) = \int_0^\infty y^2 f_Y(y)dy - \mu_Y^2 = \int_0^\infty y^2(ye^{-y})dy - \mu_Y^2 = \int y^3e^{-y}dy - \mu_Y^2$$

Using integration by parts multiple times

…

$$= 6 - 2^2 = 2$$

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{1}{\sqrt{1\times 2}} = \frac{\sqrt{2}}{2}$$

5-46. If X and Y are independent, then $f_{XY}(x,y) = f_X(x)f_Y(y)$ and the range of (X, Y) is rectangular. Therefore,

$$E(XY) = \iint xyf_X(x)f_Y(y)dxdy = \int xf_X(x)dx\int yf_Y(y)dy = E(X)E(Y)$$

hence $\sigma_{XY} = 0$

5-47. Suppose the correlation between X and Y is ρ. For constants a, b, c, and d, what is the correlation between the random variables U = aX+b and V = cY+d?

Now, E(U) = a E(X) + b and E(V) = c E(Y) + d.
Also, U - E(U) = a[X - E(X)] and V - E(V) = c[Y - E(Y)]. Then,

$$\sigma_{UV} = E\{[U - E(U)][V - E(V)]\} = acE\{[X - E(X)][Y - E(Y)]\} = ac\sigma_{XY}$$

Also, $\sigma_U^2 = E[U - E(U)]^2 = a^2 E[X - E(X)]^2 = a^2\sigma_X^2$ *and* $\sigma_V^2 = c^2\sigma_Y^2$. Then,

$$\rho_{UV} = \frac{ac\sigma_{XY}}{\sqrt{a^2\sigma_X^2}\sqrt{c^2\sigma_Y^2}} = \begin{cases} \rho_{XY} & \text{if a and c are of the same sign} \\ -\rho_{XY} & \text{if a and c differ in sign} \end{cases}$$

Section 5-3

5-48. a) board failures caused by assembly defects with probability $p_1 = 0.5$
board failures caused by electrical components with probability $p_2 = 0.3$
board failures caused by mechanical defects with probability $p_3 = 0.2$

$$P(X = 5, Y = 3, Z = 2) = \frac{10!}{5!3!2!}0.5^5 0.3^3 0.2^2 = 0.0851$$

b) Because X is binomial, $P(X = 8) = \binom{10}{8}0.5^8 0.5^2 = 0.0439$

c) $P(X = 8 \mid Y = 1) = \dfrac{P(X = 8, Y = 1)}{P(Y = 1)}$. Now, because x+y+z = 10,

P(X=8, Y=1) = P(X=8, Y=1, Z=1) = $\dfrac{10!}{8!1!1!}0.5^8 0.3^1 0.2^1 = 0.0211$

$$P(Y = 1) = \binom{10}{1}0.3^1 0.7^9 = 0.1211$$

$$P(X = 8 \mid Y = 1) = \frac{P(X = 8, Y = 1)}{P(Y = 1)} = \frac{0.0211}{0.1211} = 0.1742$$

d) $P(X \ge 8 \mid Y = 1) = \dfrac{P(X = 8, Y = 1)}{P(Y = 1)} + \dfrac{P(X = 9, Y = 1)}{P(Y = 1)}$. Now, because x+y+z = 10,

P(X=8, Y=1) = P(X=8, Y=1, Z=1) = $\dfrac{10!}{8!1!1!}0.5^8 0.3^1 0.2^1 = 0.0211$

P(X=9, Y=1) = P(X=9, Y=1, Z=0) = $\dfrac{10!}{9!1!0!}0.5^9 0.3^1 0.2^0 = 0.0059$

$$P(Y = 1) = \binom{10}{1}0.3^1 0.7^9 = 0.1211$$

$$P(X \ge 8 \mid Y = 1) = \frac{P(X = 8, Y = 1)}{P(Y = 1)} + \frac{P(X = 9, Y = 1)}{P(Y = 1)} = \frac{0.0211}{0.1211} + \frac{0.0059}{0.1211} = 0.2230$$

e) $P(X = 7, Y = 1 \mid Z = 2) = \dfrac{P(X = 7, Y = 1, Z = 2)}{P(Z = 2)}$

P(X=7, Y=1, Z=2) = $\dfrac{10!}{7!1!2!}0.5^7 0.3^1 0.2^2 = 0.0338$

$$P(Z = 2) = \binom{10}{2}0.2^2 0.8^8 = 0.3020$$

$$P(X = 7, Y = 1 \mid Z = 2) = \frac{P(X = 7, Y = 1, Z = 2)}{P(Z = 2)} = \frac{0.0338}{0.3020} = 0.1119$$

5-49. a) percentage of slabs classified as high with probability $p_1 = 0.05$
percentage of slabs classified as medium with probability $p_2 = 0.85$

percentage of slabs classified as low with probability $p_3 = 0.10$

b) X is the number of voids independently classified as high $X \geq 0$
Y is the number of voids independently classified as medium $Y \geq 0$
Z is the number of with a low number of voids and $Z \geq 0$ and $X+Y+Z = 20$

c) p_1 is the percentage of slabs classified as high.

d) $E(X)=np_1 = 20(0.05) = 1$ and $V(X)=np_1(1-p_1)= 20(0.05)(0.95) = 0.95$

e) $P(X=1, Y=17, Z=3)=0$ because the values (1, 17, 3) are not in the range of (X,Y,Z). The sum 1 + 17 + 3 > 20.

f)
$$P(X \leq 1, Y=17, Z=3) = P(X=0, Y=17, Z=3) + P(X=1, Y=17, Z=3)$$
$$= \frac{20!}{0!17!3!}0.05^0 0.85^{17} 0.10^3 + 0 = 0.07195$$
The second probability is zero because the point (1, 17, 3) is not in the range of (X, Y, Z).

g) Because X has binomial distribution,
$$P(X \leq 1) = \binom{20}{0}0.05^0 0.95^{20} + \binom{20}{1}0.05^1 0.95^{19} = 0.7358$$

h) Because X has a binomial distribution
E(Y) = np = 20(0.85) = 17

i) The probability is 0 because x + y + z > 20

j) $P(X=2 \mid Y=17) = \dfrac{P(X=2, Y=17)}{P(Y=17)}$. Now, because x + y + z = 20,

P(X=2, Y=17) = P(X=2, Y=17, Z=1) = $\dfrac{20!}{2!17!1!}0.05^2 0.85^{17} 0.10^1 = 0.0540$

$$P(X=2 \mid Y=17) = \frac{P(X=2, Y=17)}{P(Y=17)} = \frac{0.0540}{0.2428} = 0.2224$$

k)
$$E(X \mid Y=17) = 0\left(\frac{P(X=0, Y=17)}{P(Y=17)}\right) + 1\left(\frac{P(X=1, Y=17)}{P(Y=17)}\right)$$
$$+ 2\left(\frac{P(X=2, Y=17)}{P(Y=17)}\right) + 3\left(\frac{P(X=3, Y=17)}{P(Y=17)}\right)$$
$$E(X \mid Y=17) = 0\left(\frac{0.07195}{0.2428}\right) + 1\left(\frac{0.1079}{0.2428}\right) + 2\left(\frac{0.05396}{0.2428}\right) + 3\left(\frac{0.008994}{0.2428}\right)$$
$$= 1$$

5-50. a) probability for the *k*th landing page = $p_k = 0.25$

$$P(W=5, X=5, Y=5, Z=5) = \frac{20!}{5!5!5!5!}0.25^5 0.25^5 0.25^5 0.25^5 = 0.0107$$

b) Because w + x + y + z = 20, $P(W=5, X=5, Y=5) = P(W=5, X=5, Y=5, Z=5)$

$$P(W=5, X=5, Y=5) = \frac{20!}{5!5!5!5!}0.25^5 0.25^5 0.25^5 0.25^5 = 0.0107$$

c) $P(W=7, X=7, Y=6 \mid Z=3) = 0$ Because the point (7, 7 , 6, 3) is not in the range of (W,X,Y,Z).

d) $P(W=7, X=7, Y=3 \mid Z=3) = \dfrac{P(W=7, X=7, Y=3, Z=3)}{P(Z=3)}$

P(W=7,X=7, Y=3, Z=3) = $\dfrac{20!}{7!7!3!3!}0.25^7 0.25^7 0.25^3 0.25^3 = 0.0024$

$$P(Z=3) = \binom{20}{3}0.25^3 0.75^{17} = 0.1339$$

$$P(W=7, X=7, Y=3 \mid Z=3) = \frac{P(W=7, X=7, Y=3, Z=3)}{P(Z=3)} = \frac{0.0024}{0.1339} = 0.0179$$

e) Because W has a binomial distribution,

$$P(W \le 2) = \binom{20}{0}0.25^0 0.75^{20} + \binom{20}{1}0.25^1 0.75^{19} + \binom{20}{2}0.25^2 0.75^{18} = 0.0913$$

f) E(W)=np_1 = 20(0.25) = 5

g) $P(W=5, X=5) = P(W=5, X=5, Y+Z=10) = \dfrac{20!}{5!5!10!}0.25^5 0.25^5 0.5^{10} = 0.0434$

h) $P(W=5 \mid X=5) = \dfrac{P(W=5, X=5)}{P(X=5)}$

and from part g) $P(W=5, X=5) = 0.0434$

$$P(X=5) = \binom{20}{5}0.25^5 0.75^{15} = 0.2023$$

$$P(W=5 \mid X=5) = \frac{P(W=5, X=5)}{P(X=5)} = \frac{0.0434}{0.2023} = 0.2145$$

5-51.

a) The probability distribution is multinomial because the result of each trial (a dropped oven) results in either a major, minor or no defect with probability 0.6, 0.3 and 0.1 respectively. Also, the trials are independent

b) Let X, Y, and Z denote the number of ovens in the sample of four with major, minor, and no defects, respectively.

$$P(X=2, Y=2, Z=0) = \frac{4!}{2!2!0!}0.6^2 0.3^2 0.1^0 = 0.1944$$

c) $P(X=0, Y=0, Z=4) = \dfrac{4!}{0!0!4!}0.6^0 0.3^0 0.1^4 = 0.0001$

d) $f_{XY}(x,y) = \sum_R f_{XYZ}(x,y,z)$ where R is the set of values for z such that x + y + z = 4. That is, R consists of the single value z = 4 – x – y and

$$f_{XY}(x,y) = \frac{4!}{x!\,y!(4-x-y)!}0.6^x 0.3^y 0.1^{4-x-y} \quad for\ x+y \le 4.$$

e) $E(X) = np_1 = 4(0.6) = 2.4$

f) $E(Y) = np_2 = 4(0.3) = 1.2$

g) $P(X = 2 \mid Y = 2) = \dfrac{P(X = 2, Y = 2)}{P(Y = 2)} = \dfrac{0.1944}{0.2646} = 0.7347$

$P(Y = 2) = \binom{4}{2} 0.3^2 0.7^4 = 0.2646$ from the binomial marginal distribution of Y

h) Not possible, $x+y+z = 4$, the probability is zero.

i) $P(X \mid Y = 2) = P(X = 0 \mid Y = 2), P(X = 1 \mid Y = 2), P(X = 2 \mid Y = 2)$

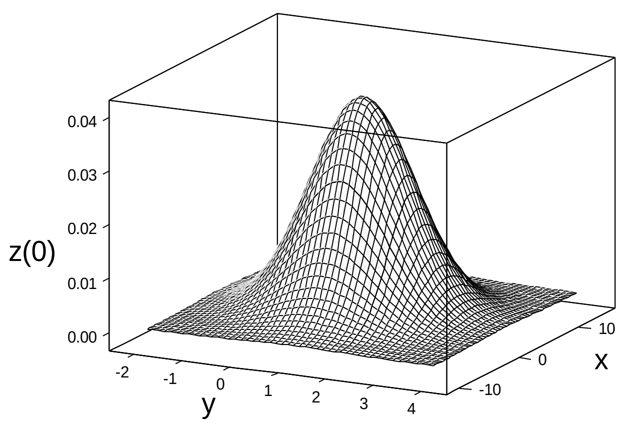
$$P(X = 0 \mid Y = 2) = \frac{P(X = 0, Y = 2)}{P(Y = 2)} = \left(\frac{4!}{0!2!2!} 0.6^0 0.3^2 0.1^2 \right) \Big/ 0.2646 = 0.0204$$

$$P(X = 1 \mid Y = 2) = \frac{P(X = 1, Y = 2)}{P(Y = 2)} = \left(\frac{4!}{1!2!1!} 0.6^1 0.3^2 0.1^1 \right) \Big/ 0.2646 = 0.2449$$

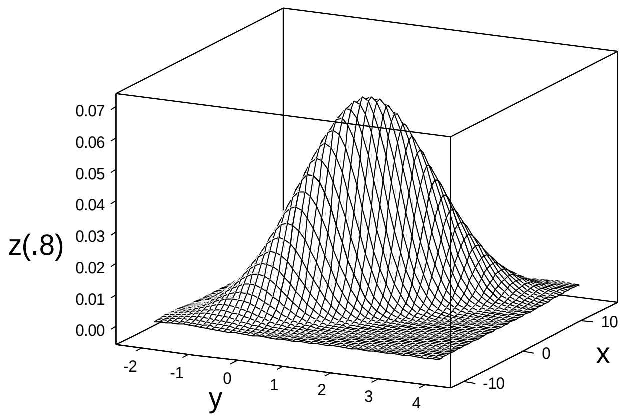
$$P(X = 2 \mid Y = 2) = \frac{P(X = 2, Y = 2)}{P(Y = 2)} = \left(\frac{4!}{2!2!0!} 0.6^2 0.3^2 0.1^0 \right) \Big/ 0.2646 = 0.7347$$

j) $E(X|Y = 2) = 0(0.0204) + 1(0.2449) + 2(0.7347) = 1.7143$

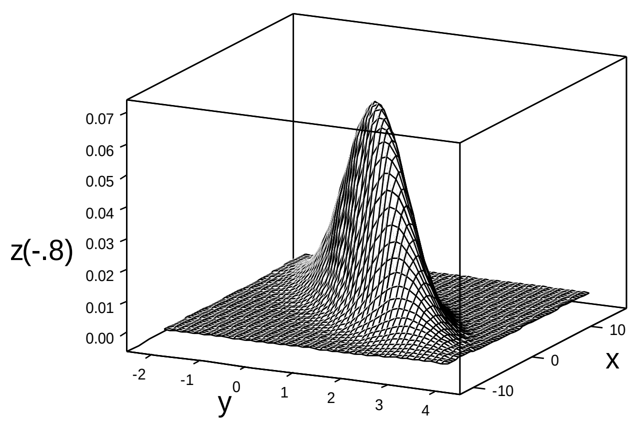
5-52. a)



b)



c)



5-53. Because $\rho = 0$ and X and Y are normally distributed, X and Y are independent. Therefore,

a) P(2.95 < X < 3.05) = $P(\frac{2.95-3}{0.04} < Z < \frac{3.05-3}{0.04})$ = 0.7887

b) P(7.60 < Y < 7.80) = $P(\frac{7.60-7.70}{0.08} < Z < \frac{7.80-7.70}{0.08})$ = 0.7887

c) P(2.95 < X < 3.05, 7.60 < Y < 7.80) = P(2.95 < X < 3.05) P(7.60 < Y < 7.80) =

$$P(\frac{2.95-3}{0.04} < Z < \frac{3.05-3}{0.04})P(\frac{7.60-7.70}{0.08} < Z < \frac{7.80-7.70}{0.08}) = 0.7887^2 = 0.6220$$

5-54. a) $\rho = cov(X,Y)/\sigma_x\sigma_y = 0.6$, cov(X,Y)= 0.6(2)(5) = 6

b) The marginal probability distribution of X is normal with mean μ_x, σ_x.

c) P(X < 116) =P(X-120 < -4) = P((X_120)/5 < -0.8) = P(Z < -0.8) = 0.21

d) The conditional probability distribution of X given Y=102 is bivariate normal distribution with mean and variance

$\mu_{X|y=102}$ = 120 – 100(0.6)(5/2) +(5/2)(0.6)(102) = 123

$\sigma^2_{X|y=102}$ = 25(1-0.36) = 16

e)P(X < 116|Y=102)=P(Z < (116-123)/4)=0.040

5-55. Because $\rho = 0$ and X and Y are normally distributed, X and Y are independent. Therefore, μ_X = 0.1 mm, σ_X=0.00031 mm, μ_Y = 0.23 mm, σ_Y=0.00017 mm

Probability X is within specification limits is

$$P(0.099535 < X < 0.100465) = P\left(\frac{0.099535-0.1}{0.00031} < Z < \frac{0.100465-0.1}{0.00031}\right)$$
$$= P(-1.5 < Z < 1.5) = P(Z < 1.5) - P(Z < -1.5) = 0.8664$$

Probability that Y is within specification limits is

$$P(0.22966 < X < 0.23034) = P\left(\frac{0.22966-0.23}{0.00017} < Z < \frac{0.23034-0.23}{0.00017}\right)$$
$$= P(-2 < Z < 2) = P(Z < 2) - P(Z < -2) = 0.9545$$

Probability that a randomly selected lamp is within specification limits is (0.8664)(0.9594) = 0.8270

5-56.

a) X_1, X_2 and X_3 are binomial random variables when considered individually, i.e., the marginal probability distributions are binomial. Their joint distribution is multinomial.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$$
$$= \frac{20!}{x_1!x_2!x_3!}(0.5)^{x_1}(0.4)^{x_2}(0.1)^{x_3}$$
$$\neq P(X_1 = x_1)P(X_2 = x_2)P(X_3 = x_3)$$
$$= \binom{20}{x_1}(0.5)^{x_1}(1-0.5)^{20-x_1}\binom{20}{x_2}(0.4)^{x_2}(1-0.4)^{20-x_2}\binom{20}{x_3}(0.1)^{x_3}(1-0.1)^{20-x_3}$$

Hence, X_1, X_2 and X_3 are not independent.

b)

$$P(X_1 = 10) = \binom{20}{10}0.5^{10}(1-0.5)^{10} \cong 0.176$$

c)

$$P(X_1 = 10, X_2 = 8, X_3 = 2)$$
$$= \frac{20!}{(10!)(8!)(2!)}(0.5)^{10}(0.4)^8(0.1)^2 = 0.0532$$

d)

$$P(X_1 = 5 \mid X_2 = 12) = \frac{P(X_1 = 5, X_2 = 12)}{P(X_2 = 12)}$$

$$= \frac{\frac{20!}{(5!)(12!)(3!)}(0.5)^5(0.4)^{12}(0.1)^3}{\binom{20}{12}(0.4)^{12}(1-0.4)^8} = 0.1042$$

e)

$$E(X_1) = np_1 = 20(0.5) = 10$$

5-57.

a)

Because $f_{Y|x}(y) \neq f_Y(y)$ for all *x* and *y* with $f_X(x) > 0$, *X* and *Y* are not independent.

b) The conditional mean of *Y* given *X* = 3 is 2(3) = 6

$$P(Y < 3 \mid X = 3) = P\left(\frac{Y-6}{2} \le \frac{3-6}{2}\right) = P(Z \le -1.5) = 0.0668$$

c) $E(Y \mid X = 3) = 2 \times 3 = 6$

d)

From the formulas for the mean and variance of a conditional normal distribution we have

$$(1-\rho^2)\sigma_Y^2 = 2$$
$$\mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(x - \mu_X) = 2x$$

From the second equation

$$\frac{\sigma_Y}{\sigma_X}\rho = 2 \text{ and } \mu_Y - \frac{\sigma_Y}{\sigma_X}\rho\mu_X = 0$$

and these can be written as

$$\sigma_Y = \frac{2\sigma_X}{\rho} \text{ and } \mu_Y = \frac{\sigma_Y}{\sigma_X}\rho\mu_X$$

From the first equation above

$$(1-\rho^2)\left(\frac{2^2\sigma_X^2}{\rho^2}\right) = 2$$

Because $\sigma_X = 1$, the previous equation can be solved for ρ to yield

$$(1-\rho^2)\left(\frac{8}{\rho^2}\right) = 2$$

and ρ = 0.894

Then $\sigma_Y = \frac{2\sigma_X}{\rho} = \frac{2\sqrt{2}}{0.894} = 3.16$ and $\sigma_Y^2 = 10$

Also, $\mu_Y = \frac{\sigma_Y}{\sigma_X}\rho\mu_X = \frac{3.46}{1}(0.816)(0) = 0$

5-58. a) By completing the square in the numerator of the exponent of the bivariate normal PDF, the joint PDF can be written as

$$f_{Y|X=x}=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{\left[\frac{1}{\sigma_Y^2}\left[(y-(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_x))\right]^2+(1-\rho^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]}{2(1-\rho^2)}}}{\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{\left[\frac{x-\mu_x}{\sigma_x}\right]^2}{2}}}$$

Also, $f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{\left[\frac{x-\mu_x}{\sigma_x}\right]^2}{2}}$ By definition,

$$f_{Y|X=x}=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{\left[\frac{1}{\sigma_Y^2}\left[(y-(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_x))\right]^2+(1-\rho^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]}{2(1-\rho^2)}}}{\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{\left[\frac{x-\mu_x}{\sigma_x}\right]^2}{2(1-\rho^2)}}}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}e^{-\frac{\frac{1}{\sigma_Y^2}\left[(y-(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_x))\right]^2}{2(1-\rho^2)}}$$

Now $f_{Y|X=x}$ is in the form of a normal distribution.

b) E(Y|X=x) = $\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$. This answer can be seen from part a). Because the PDF is in the form of a normal distribution, then the mean can be obtained from the exponent.

c) V(Y|X=x) = $\sigma_y^2(1-\rho^2)$. This answer can be seen from part a). Because the PDF is in the form of a normal distribution, then the variance can be obtained from the exponent.

5-59.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X{}^2}+\frac{(y-\mu_Y)^2}{\sigma_Y{}^2}\right]}\right]dxdy=$$

$$\int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X{}^2}\right]}\right]dx\int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{1}{2}\left[\frac{(y-\mu_Y)^2}{\sigma_Y{}^2}\right]}\right]dy$$

and each of the last two integrals is recognized as the integral of a normal probability density function from $-\infty$ to ∞. That is, each integral equals one. Because $f_{XY}(x,y)=f(x)f(y)$, X and Y are independent.

5-60. Let $f_{XY}(x,y)=\dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2-\frac{2\rho(X-\mu_X)(Y-\mu_X)}{\sigma_X\sigma_Y}+\left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right]}{2(1-\rho^2)}}$

Completing the square in the numerator of the exponent we get:

$$\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^2-\frac{2\rho(X-\mu_X)(Y-\mu_X)}{\sigma_X\sigma_Y}+\left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right]=\left[\left(\frac{Y-\mu_Y}{\sigma_Y}\right)-\rho\left(\frac{X-\mu_X}{\sigma_X}\right)\right]^2+(1-\rho^2)\left(\frac{X-\mu_X}{\sigma_X}\right)^2$$

But,

$$\left(\frac{Y-\mu_Y}{\sigma_Y}\right)-\rho\left(\frac{X-\mu_X}{\sigma_X}\right)=\frac{1}{\sigma_Y}\left[(Y-\mu_Y)-\rho\frac{\sigma_Y}{\sigma_X}(X-\mu_x)\right]=\frac{1}{\sigma_Y}\left[(Y-(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(X-\mu_x))\right]$$

Substituting into $f_{XY}(x,y)$, we get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{XY}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-\frac{\left[\frac{1}{\sigma_Y^2}\left[y-(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_x))\right]^2+(1-\rho^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]}{2(1-\rho^2)}}dydx$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}dx\times\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}e^{-\left[\frac{\left(y-(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x))\right)^2}{2\sigma_x^2(1-\rho^2)}\right]}dy$$

The integrand in the second integral above is in the form of a normally distributed random variable. By definition of the integral over this function, the second integral is equal to 1:

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}dx\times\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}e^{-\left[\frac{\left(y-(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x))\right)^2}{2\sigma_x^2(1-\rho^2)}\right]}dy$$

$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}dx\times 1$$

The remaining integral is also the integral of a normally distributed random variable and therefore, it also integrates to 1, by definition. Therefore,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{XY}(x,y)=1$$

5-61.

$$f_X(x) = \int_{-\infty}^{\infty} \left[\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\frac{-0.5}{1-\rho^2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]} \right] dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} e^{\frac{-0.5}{1-\rho^2}\frac{(x-\mu_X)^2}{\sigma_X^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} e^{\frac{-0.5}{1-\rho^2}\left\{\left[\frac{(y-\mu_Y)}{\sigma_Y} - \frac{\rho(x-\mu_X)}{\sigma_X}\right]^2 - \left[\frac{\rho(x-\mu_X)}{\sigma_X}\right]^2\right\}} dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-0.5\frac{(x-\mu_X)^2}{\sigma_X^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} e^{\frac{-0.5}{1-\rho^2}\left[\frac{(y-\mu_Y)}{\sigma_Y} - \frac{\rho(x-\mu_X)}{\sigma_X}\right]^2} dy$$

The last integral is recognized as the integral of a normal probability density with mean $\mu_Y + \frac{\sigma_Y\rho(x-\mu_X)}{\sigma_X}$ and variance $\sigma_Y{}^2(1-\rho^2)$. Therefore, the last integral equals one and the requested result is obtained.

Section 5-4

5-62. a) E(2X + 3Y) = 2(0) + 3(10) = 30
b) V(2X + 3Y) = 4V(X) + 9V(Y) = 97
c) 2X + 3Y is normally distributed with mean 30 and variance 97. Therefore,

$$P(2X+3Y<30) = P(Z<\tfrac{30-30}{\sqrt{97}}) = P(Z<0) = 0.5$$

d) $P(2X+3Y<40) = P(Z<\frac{40-30}{\sqrt{97}}) = P(Z<1.02) = 0.8461$

5-63.
a) E(3X+2Y) = 3(2) + 2(6) = 18
b) V(3X+2Y) = 9(5) + 4(8) = 77
c) 3X+2Y ~ N(18, 77), P(3X+2Y < 18) = P(Z < (18-18)/$77^{0.5}$) = 0.5
d) P(3X+2Y < 28) = P(Z < (28-18)/$77^{0.5}$) = P(Z < 1.1396) = 0.873

5-64. Y = 10X and E(Y) =10E(X) = 50mm. V(Y) = 10^2V(X) = 25mm^2

5-65. a) Let T denote the total thickness. Then, T = X + Y and E(T) = 4 mm
V(T) = $0.1^2 + 0.1^2 = 0.02$ mm^2 and $\sigma_T = 0.1414$ mm.
b)

$$P(T>4.3) = P\left(Z > \frac{4.3-4}{0.1414}\right) = P(Z>2.12)$$
$$= 1 - P(Z<2.12) = 1 - 0.983 = 0.0170$$

5-66. a) X: time of wheel throwing. X~N(40,4)
Y: time of wheel firing. Y~N(60,9)
X+Y ~ N(100, 13)
P(X+Y≤ 95) = P(Z<(95-100)/$13^{0.5}$) = P(Z<-1.387) = 0.083

b) P(X+Y>110) = 1- P(Z<(110-100)/ $13^{0.5}$) =1-P(Z<2.774) = 1-0.9972 =0.0028

5-67. a) *X*~N(0.1, 0.00031) and *Y*~N(0.23, 0.00017) Let T denote the total thickness.

Then, T = X + Y and E(T) = 0.33 mm,
V(T) = $0.00031^2 + 0.00017^2 = 1.25x10^{-7}$ mm², and $\sigma_T = 0.000354$ mm.

$$P(T < 0.2337) = P\left(Z < \frac{0.2337 - 0.33}{0.000354}\right) = P(Z < -272) \cong 0$$

b)

$$P(T > 0.2405) = P\left(Z > \frac{0.2405 - 0.33}{0.000345}\right) = P(Z > -253) = 1 - P(Z < 253) \cong 1$$

5-68. Let D denote the width of the casing minus the width of the door. Then, D is normally distributed.

a) E(D) = 1/8 V(D) = $(\frac{1}{8})^2 + (\frac{1}{16})^2 = \frac{5}{256}$ $\sigma_T = \sqrt{5/256} = 0.1398$

b) $P(D > \frac{1}{4}) = P(Z > \frac{\frac{1}{4} - \frac{1}{8}}{\sqrt{5/256}}) = P(Z > 0.89) = 0.187$

c) $P(D < 0) = P(Z < \frac{0 - \frac{1}{8}}{\sqrt{5/256}}) = P(Z < -0.89) = 0.187$

5-69. X = time of ACL reconstruction surgery for high-volume hospitals.
X ~ N(129,196)
$E(X_1+X_2+\ldots+X_{10}) = 10(129) = 1290$
$V(X_1+X_2+\ldots+X_{10}) = 100(196) = 19600$

5-70. a) Let $\overline{X}$ denote the average fill-volume of 100 cans. $\sigma_{\overline{X}} = \sqrt{0.5^2/100} = 0.05$

b) E($\overline{X}$) = 12.1 and $P(\overline{X} < 12) = P\left(Z < \frac{12 - 12.1}{0.05}\right) = P(Z < -2) = 0.023$

c) P($\overline{X}$ < 12) = 0.005 implies that $P\left(Z < \frac{12 - \mu}{0.05}\right) = 0.005.$

Then $\frac{12-\mu}{0.05}$ = -2.58 and $\mu = 12.129$

d) P($\overline{X}$ < 12) = 0.005 implies that $P\left(Z < \frac{12 - 12.1}{\sigma / \sqrt{100}}\right) = 0.005.$

Then $\frac{12-12.1}{\sigma/\sqrt{100}}$ = -2.58 and $\sigma = 0.388$

e) P($\overline{X}$ < 12) = 0.01 implies that $P\left(Z < \frac{12 - 12.1}{0.5 / \sqrt{n}}\right) = 0.01.$

Then $\frac{12-12.1}{0.5/\sqrt{n}}$ = -2.33 and $n = 135.72 \cong 136$.

5-71. Let $\overline{X}$ denote the average thickness of 10 wafers. Then, E($\overline{X}$) = 10 and V($\overline{X}$) = 0.1.

a) $P(9 < \overline{X} < 11) = P(\frac{9-10}{\sqrt{0.1}} < Z < \frac{11-10}{\sqrt{0.1}}) = P(-3.16 < Z < 3.16) = 0.998$.

The answer is 1 – 0.998 = 0.002

b) $P(\overline{X} > 11) = 0.01$ *and* $\sigma_{\overline{X}} = \frac{1}{\sqrt{n}}$.

Therefore, $P(\overline{X} > 11) = P(Z > \frac{11-10}{1/\sqrt{n}}) = 0.01$, $\frac{11-10}{1/\sqrt{n}}$ = 2.33 and n = 5.43 which is rounded up to 6.

c) $P(\overline{X} > 11) = 0.0005 \;\; and \;\; \sigma_{\overline{X}} = \sigma/\sqrt{10}$.

Therefore, $P(\overline{X} > 11) = P(Z > \frac{11-10}{\sigma/\sqrt{10}}) = 0.0005$, $\frac{11-10}{\sigma/\sqrt{10}}$ = 3.29

$\sigma = \sqrt{10} / 3.29 = 0.9612$

5-72. X~N(160, 900)

a) Let Y = $X_1 + X_2 + ... + X_{25}$, E(Y) = 25E(X) = 4000, V(Y) = 25^2(900) = 22500

P(Y > 4300) =

$$P(Y > 4300) = P\left(Z > \frac{4300-4000}{\sqrt{22500}}\right) = P(Z > 2) = 1 - 0.9773 = 0.0227$$

b) P(Y > x) = 0.0001 implies that $P\left(Z > \frac{x-4000}{\sqrt{22500}}\right) = 0.0001$

Then $\frac{x-4000}{150}$ = 3.72 and $x = 4558$

5-73. W = weights of a part and E = measurement error.

W~ N(μ_w, σ_w^2), E ~ N(0, σ_e^2),W+E ~ N(μ_w, $\sigma_w^2+\sigma_e^2$) .

W_{sp} = weight of the specification P

a) P(W > $\mu_w + 3\sigma_w$) + P(W < $\mu_w - 3\sigma_w$) = P(Z > 3) + P(Z < -3) = 0.0027

b) P(W+E > $\mu_w + 3\sigma_w$) + P(W+E < $\mu_w - 3\sigma_w$)

= P (Z > $3\sigma_w / (\sigma_w^2+\sigma_e^2)^{1/2}$) + P (Z < $-3\sigma_w / (\sigma_w^2+\sigma_e^2)^{1/2}$)

Because $\sigma_e^2 = 0.5\sigma_w^2$ the probability is

= P (Z > $3\sigma_w / (1.5\sigma_w^2)^{1/2}$) + P (Z < $-3\sigma_w / (1.5\sigma_w^2)^{1/2}$)

= P (Z > 2.45) + P (Z < -2.45) = 2(0.0072) = 0.014

No.

c) P(E + $\mu_w + 2\sigma_w > \mu_w + 3\sigma_w$) = P(E > σ_w) = P(Z > $\sigma_w/(0.5\sigma_w^2)^{1/2}$) = P(Z > 1.41) = 0.079

Also, P(E + $\mu_w + 2\sigma_w < \mu_w - 3\sigma_w$) = P(E < $-5\sigma_w$) = P(Z < $-5\sigma_w/(0.5\sigma_w^2)^{1/2}$) = P(Z < -7.07) ≈ 0

5-74. D = A - B - C

a) E(D) = 10 - 2 - 2 = 6 mm

$$V(D) = 0.1^2 + 0.05^2 + 0.05^2 = 0.015$$

$$\sigma_D = 0.1225$$

b) P(D < 5.9) = P(Z < $\frac{5.9-6}{0.1225}$) = P(Z < -0.82) = 0.206

5-75.

$$V(Y) = V(2X_1 + 2X_2)$$
$$= 2^2 V(X_1) + 2^2 V(X_2) + 2 \times 2 \times 2 \times Cov(X_1, X_2)$$
$$= 4(0.1)^2 + 4(0.2)^2 + 8(0.02) = 0.36$$

The positive covariance increases the variance of the perimeter. In the example, the random variables were assumed to be independent.

5-76. $Y = X_1 + X_2 + X_3$

a)

$$E(Y) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3)$$
$$= \frac{3+7}{2} + \frac{2+5}{2} + \frac{4+10}{2} = 15.5$$
$$V(Y) = V(X_1) + V(X_2) + V(X_3)$$
$$= \frac{(7-3)^2}{12} + \frac{(5-2)^2}{12} + \frac{(10-4)^2}{12} = 5.083$$

b)

$$E(Y) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3)$$
$$= \frac{3+7}{2} + \frac{2+5}{2} + \frac{4+10}{2} = 15.5$$
$$V(Y) = V(X_1) + V(X_2) + V(X_3) + 2\big(Cov(X_1, X_2) + Cov(X_1, X_3) + Cov(X_2, X_3)\big)$$
$$= \frac{61}{12} + 2 \times [(-0.5) + (-0.5) + (-0.5)] = 2.083$$

c) The covariance does not have an impact on $E(Y)$. However, negative covariance results in a decrease in $V(Y)$.

5-77.

Let X_i be the demand in month i.

a)

$$Z = X_1 + X_2 + ... + X_{12}$$
$$E(Z) = E(X_1 + X_2 + ... + X_{12}) = E(X_1) + E(X_2) + ... + E(X_{12}) = 12 \times (1.1) = 13.2$$
$$V(Z) = V(X_1) + V(X_2) + ... + V(X_{12}) = 12 \times (0.3)^2 = 1.08$$

Z is normally distributed with a mean of 13.2 and a standard deviation of $\sqrt{1.08}$ millions of doses.

b)

$$P(Z < 13.2) = P\left(\frac{Z - 13.2}{\sqrt{1.08}} < 0\right) = 0.5$$

c)

$$P(11 < Z < 15) = P\left(\frac{11 - 13.2}{\sqrt{1.08}} < Z < \frac{15 - 13.2}{\sqrt{1.08}}\right) = 0.9412$$

d)

$$P(Z < c) = 0.99$$
$$P\left(\frac{Z - 13.2}{\sqrt{1.08}} < \frac{c - 13.2}{\sqrt{1.08}}\right) = 0.99$$
$$\Rightarrow \frac{c - 13.2}{\sqrt{1.08}} = 2.33 \Rightarrow c = 15.62$$

5-78.

Let Y be the rate of return for the entire investment after one year.

a)

$$E(Y) = c_1 E(X_1) + c_2 E(X_2) + c_3 E(X_3)$$

where $c_1 = \frac{2500}{10000} = 0.25, c_2 = \frac{3000}{10000} = 0.30,$ and $c_3 = \frac{4500}{10000} = 0.45$

$$E(Y) = 0.25 \times 0.12 + 0.30 \times 0.04 + 0.45 \times 0.07 = 0.0735$$
$$V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + c_3^2 V(X_3)$$
$$= (0.25)^2 (0.14)^2 + (0.30)^2 (0.02)^2 + (0.45)^2 (0.08)^2 = 0.002557 \approx 0.003$$

b) We are given $Cov(X_2, X_3) = -0.005$

$$E(Y) = 0.25 \times 0.12 + 0.30 \times 0.04 + 0.45 \times 0.07 = 0.0735$$
$$V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + c_3^2 V(X_1) + 2c_2 c_3 Cov(X_2, X_3)$$
$$= (0.25)^2 (0.14)^2 + (0.30)^2 (0.02)^2 + (0.45)^2 (0.08)^2 + 2 \times 0.30 \times 0.45 \times (-0.005)$$
$$= 0.001207 \cong 0.001$$

c) The covariance does not have an effect on the mean rate of return for the entire investment. However, negative covariance between two assets decreases the variance of the rate of return for the entire investment which reduces the risk of the investment.

Section 5-5

5-79. $f_Y(y) = \frac{1}{4}$ at y = 3, 5, 7, 9

5-80. Because $X \geq 0$, the transformation is one-to-one; that is, $y = x^2$ and $x = \sqrt{y}$.

Now, $f_Y(y) = f_X(\sqrt{y}) = \binom{3}{\sqrt{y}} p^{\sqrt{y}} (1-p)^{3-\sqrt{y}}$ for y = 0, 1, 4, 9.

If p = 0.25, $f_Y(y) = \binom{3}{\sqrt{y}} (0.25)^{\sqrt{y}} (0.75)^{3-\sqrt{y}}$ for y = 0, 1, 4, 9.

5-81. a) $f_Y(y) = f_X\left(\frac{y-10}{2}\right)\left(\frac{1}{2}\right) = \frac{y-10}{72}$ for $10 \leq y \leq 22$

b) $E(Y) = \int_{10}^{22} \frac{y^2 - 10y}{72} dy = \frac{1}{72}\left(\frac{y^3}{3} - \frac{10y^2}{2}\right)\Big|_{10}^{22} = 18$

5-82. Because y = -2 ln x, $e^{-\frac{y}{2}} = x$. Then, $f_Y(y) = f_X(e^{\frac{-y}{2}})\left|-\frac{1}{2}e^{\frac{-y}{2}}\right| = \frac{1}{2}e^{\frac{-y}{2}}$ for $0 \leq e^{\frac{-y}{2}} \leq 1$ or $y \geq 0$, which is an exponential distribution with λ = 1/2 (which equals a chi-square distribution with k = 2 degrees of freedom).

5-83. a) If $y = x^2$, then $x = \sqrt{y}$ for $x \geq 0$ and $y \geq 0$. Thus, $f_Y(y) = f_X(\sqrt{y})\frac{1}{2}y^{-\frac{1}{2}} = \frac{e^{-\sqrt{y}}}{2\sqrt{y}}$ for y > 0.

b) If $y = x^{1/2}$, then $x = y^2$ for $x \geq 0$ and $y \geq 0$. Thus, $f_Y(y) = f_X(y^2)2y = 2ye^{-y^2}$ for y > 0.

c) If y = ln x, then x = e^y for $x \ge 0$. Thus, $f_Y(y) = f_X(e^y)e^y = e^y e^{-e^y} = e^{y-e^y}$ for $-\infty < y < \infty$.

5-84. a) Now, $\int_0^\infty av^2 e^{-bv} dv$ must equal one. Let *u* = bv, then $1 = a\int_0^\infty (\frac{u}{b})^2 e^{-u} \frac{du}{b} = \frac{a}{b^3}\int_0^\infty u^2 e^{-u} du.$

From the definition of the gamma function the last expression is $\frac{a}{b^3}\Gamma(3) = \frac{2a}{b^3}$. Therefore,

$a = \frac{b^3}{2}$.

b) If $w = \frac{mv^2}{2}$, then $v = \sqrt{\frac{2w}{m}}$ for $v \ge 0, \quad w \ge 0$.

$$f_W(w) = f_V\left(\sqrt{\tfrac{2w}{m}}\right)\frac{dv}{dw} = \frac{b^3 2w}{2m} e^{-b\sqrt{\frac{2w}{m}}} (2mw)^{-1/2}$$

$$= \frac{b^3 m^{-3/2}}{\sqrt{2}} w^{1/2} e^{-b\sqrt{\frac{2w}{m}}}$$

for $w \ge 0$.

5-85. If $y = e^x$, then x = ln y for $1 \le x \le 2$ and $e^1 \le y \le e^2$. Thus, $f_Y(y) = f_X(\ln y)\frac{1}{y} = \frac{1}{y}$

for $1 \le \ln y \le 2$. That is, $f_Y(y) = \frac{1}{y}$ for $e \le y \le e^2$.

5-86. If y = $(x-2)^2$, then x = $2-\sqrt{y}$ for $0 \le x \le 2$ and x = $2+\sqrt{y}$ for $2 \le x \le 4$. Thus,

$$f_Y(y) = f_X(2-\sqrt{y})\,|-\tfrac{1}{2}y^{-1/2}| + f_X(2+\sqrt{y})\,|\tfrac{1}{2}y^{-1/2}|$$

$$= \frac{2-\sqrt{y}}{16\sqrt{y}} + \frac{2+\sqrt{y}}{16\sqrt{y}}$$

$$= \left(\tfrac{1}{4}\right)y^{-1/2} \text{ for } 0 \le y \le 4$$

5-87.

$$r = \left(r_1^2 + r_2^2 + r_1 r_2(\cos\theta_1 - \cos\theta_2)\right)^{0.5}$$

$$\theta_2 = \arccos\left(\cos\theta_1 - \frac{r^2 - r_1^2 - r_2^2}{r_1 r_2}\right)$$

Then, $f(r) = f(\theta_2)|J|$, where

$$J = \frac{d}{dr}\arccos\left(\cos\theta_1 - \frac{r^2 - r_1^2 - r_2^2}{r_1 r_2}\right)$$

$$= \frac{-1}{\sqrt{1-\left(\cos\theta_1 - \frac{r^2 - r_1^2 - r_2^2}{r_1 r_2}\right)^2}}\frac{-2r}{r_1 r_2} = \frac{2r}{\sqrt{1-\left(\cos\theta_1 - \frac{r^2 - r_1^2 - r_2^2}{r_1 r_2}\right)^2}}\frac{1}{r_1 r_2}$$

and $f(\theta_2) = \frac{1}{10}$

5-88.

$$Y = e^W$$

$$W = \ln Y$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\omega} e^{\frac{-(\ln y - \theta)^2}{2\omega^2}} |J|$$

where $J = \frac{d}{dy}W = \frac{d}{dy}\ln y = \frac{1}{y}$

Substituting J in $f_Y(y)$ and rearranging gives us the lognormal pdf.

$$f_Y(y) = \frac{1}{y\omega\sqrt{2\pi}} e^{-\frac{(\ln y - \theta)^2}{2\omega^2}}$$

5-89.

$$T = 0.004N^2 \Rightarrow N = \sqrt{\frac{T}{0.004}}$$

Then,

$$f_T(t) = \frac{1}{10000} e^{\frac{-1}{10000}\sqrt{\frac{t}{0.004}}} |J|$$

where

$$J = \frac{d}{dt}N = \frac{d}{dt}\sqrt{\frac{t}{0.004}} = \frac{1}{2\sqrt{0.004t}}$$

5-90.

Here X is lognormal with $\theta = 5.2933$, and $\omega^2 = 0.00995$

$$Y = X^4$$

a) $E(X) = e^{\theta + \frac{\omega^2}{2}} \cong 200$ $\qquad V(X) = e^{2\theta + \omega^2}(e^{\omega^2} - 1) \cong 400$

b) $Y = X^4$

$X = \sqrt[4]{Y}$

Then,

$$f_Y(y) = \frac{1}{\sqrt[4]{y}\omega\sqrt{2\pi}} e^{\frac{-(\ln(\sqrt[4]{y}) - \theta)^2}{2\omega^2}} |J|$$

where $J = \frac{1}{4}y^{\frac{-3}{4}}$ Substituting J into $f_Y(y)$, we have

$$f_Y(y) = \frac{1}{y(4\omega)\sqrt{2\pi}} e^{\frac{-(\ln(y)-4\theta)^2}{32\omega^2}}$$

Letting $\omega' = 4\omega$, and $\theta' = 4\theta$, we have $f_Y(y) = \dfrac{1}{y\omega'\sqrt{2\pi}} e^{\frac{-(\ln(y)-\theta')^2}{2(\omega')^2}}$

Hence, Y is lognormal with $\theta' = 21.1732$, and $\omega'^2 = 0.1592$

c) $E(Y) = e^{\theta' + \frac{\omega'^2}{2}} = 1{,}698{,}141{,}067.5$ $\quad V(Y) = e^{2\theta' + \omega'^2}(e^{\omega'^2} - 1) = 4.976 \times 10^{17}$

d) $\sqrt[4]{E(Y)} \cong 203$

e) $\sqrt[4]{E(Y)} > E(X)$ The normalized power is slightly greater than the mean power in this example.

<u>Section 5-6</u>

5-91.

a) $M_X(t) = E(e^{tX}) = \sum_{x=1}^{m} \frac{1}{m} e^{tx} = \frac{1}{m}\sum_{x=0}^{m-1} e^{t(x+1)} = \frac{e^t}{m}\sum_{x=0}^{m-1} e^{tx} = \frac{e^t(1-e^{tm})}{m(1-e^t)}$

b) $E(X) = \mu = \mu_1' = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \frac{m+1}{2}$

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = \mu_2' - \mu^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2 = \frac{m^2-1}{12}$$

5-92.

a) $M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda}(e^{\lambda e^t}) = e^{\lambda(e^t-1)}$

b) $E(X) = \mu = \mu_1' = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \lambda e^t\left(e^{\lambda(e^t-1)}\right)\Big|_{t=0} = \lambda$

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = \mu_2' - \mu^2$$

$$\mu_2' = \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \left[\lambda e^t\left(e^{\lambda(e^t-1)}\right) + \left(\lambda e^t\right)^2\left(e^{\lambda(e^t-1)}\right)\right]_{t=0} = \lambda + \lambda^2$$

$$\sigma^2 = \mu_2' - \mu^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

5-93. $M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} p(1-p)^{x-1} = \frac{p}{1-p}\sum_{x=1}^{\infty}(e^t(1-p))^x$

$$= \frac{p}{1-p}\left(\frac{e^t(1-p)}{1-e^t(1-p)}\right) = \frac{pe^t}{1-(1-p)e^t}$$

$$E(X) = \mu = \mu_1' = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \left.\frac{pe^t\left(1-(1-p)e^t\right) - pe^t\left(-(1-p)e^t\right)}{\left(1-(1-p)e^t\right)^2}\right|_{t=0}$$

$$= \left.\frac{pe^t}{\left(1-(1-p)e^t\right)^2}\right|_{t=0} = \frac{p}{p^2} = \frac{1}{p}$$

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = \mu_2' - \mu^2$$

$$\mu_2' = \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \left.\frac{pe^t\left(1-(1-p)e^t\right)^2 - pe^t(2)\left(1-(1-p)e^t\right)\left(-(1-p)e^t\right)}{\left(1-(1-p)e^t\right)^4}\right|_{t=0}$$

$$= \left.\frac{pe^t\left(1+(1-p)e^t\right)}{\left(1-(1-p)e^t\right)^3}\right|_{t=0} = \frac{2-p}{p^2}$$

$$\sigma^2 = \mu_2' - \mu^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

5-94. $M_Y(t) = M_{X_1}(t).M_{X_2}(t) = \left((1-2t)^{-k_1/2}\right)\left((1-2t)^{-k_2/2}\right) = (1-2t)^{-(k_1+k_2)/2}$

As a result, Y is a chi-squared random variable with $k_1 + k_2$ degrees of freedom.

5-95.

a) $M_X(t) = E(e^{tX}) = \int_{x=0}^{\infty} e^{tx}(4xe^{-2x})dx = 4\int_{x=0}^{\infty} xe^{(t-2)x}dx$

Using integration by parts, we have:

$$M_X(t) = 4\lim_{c\to\infty}\left[\frac{xe^{(t-2)x}}{t-2} - \frac{e^{(t-2)x}}{(t-2)^2}\right]_0^c = 4\lim_{c\to\infty}\left[\left(\frac{x}{t-2} - \frac{1}{(t-2)^2}\right)e^{(t-2)x}\right]_0^c = \frac{4}{(t-2)^2}$$

b) $E(X) = \mu = \mu_1' = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \left.\frac{-8}{(t-2)^3}\right|_{t=0} = 1$

$$\mu_2' = \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \left.\frac{24}{(t-2)^4}\right|_{t=0} = 1.5$$

$$V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = \mu_2' - \mu^2 = 1.5 - 1 = 0.5$$

5-96.

a) $M_X(t) = E(e^{tX}) = \int_{\alpha}^{\beta} e^{tx} \frac{1}{\beta-\alpha} dx = \frac{1}{\beta-\alpha}\left(\frac{e^{tx}}{t}\right)\Bigg|_{x=\alpha}^{x=\beta} = \frac{1}{\beta-\alpha}\left(\frac{e^{t\beta}}{t} - \frac{e^{t\alpha}}{t}\right) = \frac{e^{t\beta}-e^{t\alpha}}{t(\beta-\alpha)}$

b) $E(X) = \mu = \mu_1' = \frac{dM_X(t)}{dt}\bigg|_{t=0}$

$$\frac{dM_X(t)}{dt} = \frac{\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right)\left(t(\beta-\alpha)\right) - \left(e^{t\beta} - e^{t\alpha}\right)(\beta-\alpha)}{t^2(\beta-\alpha)^2} = \frac{t\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right) - e^{t\beta} + e^{t\alpha}}{t^2(\beta-\alpha)}$$

$\frac{dM_X(t)}{dt}$ is undefined at $t = 0$ since there is t^2 in the denominator. Indeed, it has an indeterminate form of $0/0$ when it is evaluated at $t = 0$. As a result, we need to use L'Hopital's rule and differentiate the numerator and denominator.

$$E(X) = \lim_{t\to 0} \frac{dM_X(t)}{dt} = \lim_{t\to 0} \frac{t\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right) - e^{t\beta} + e^{t\alpha}}{t^2(\beta-\alpha)}$$

$$= \lim_{t\to 0} \frac{\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right) + t\left(\beta^2 e^{t\beta} - \alpha^2 e^{t\alpha}\right) - \beta e^{t\beta} + \alpha e^{t\alpha}}{2t(\beta-\alpha)}$$

$$= \lim_{t\to 0} \frac{\beta^2 e^{t\beta} - \alpha^2 e^{t\alpha}}{2(\beta-\alpha)} = \frac{\beta^2 - \alpha^2}{2(\beta-\alpha)} = \frac{\alpha+\beta}{2}$$

$$\mu_2' = \frac{d^2M_X(t)}{dt^2}\bigg|_{t=0}$$

$$\frac{d^2M_X(t)}{dt^2} = \frac{t^3\left(\beta^2 e^{t\beta} - \alpha^2 e^{t\alpha}\right)(\beta-\alpha) - 2t(\beta-\alpha)\left(t\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right) - e^{t\beta} + e^{t\alpha}\right)}{t^4(\beta-\alpha)^2}$$

$$= \frac{t^2\left(\beta^2 e^{t\beta} - \alpha^2 e^{t\alpha}\right) - 2t\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right) + 2e^{t\beta} - 2e^{t\alpha}}{t^3(\beta-\alpha)}$$

$\frac{d^2M_X(t)}{dt^2}$ has the same indefinite form of $0/0$ when it is evaluated at $t = 0$. We need to use L'Hopital's rule again.

$$\mu_2' = \lim_{t\to 0} \frac{t^2\left(\beta^2 e^{t\beta} - \alpha^2 e^{t\alpha}\right) - 2t\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right) + 2e^{t\beta} - 2e^{t\alpha}}{t^3(\beta-\alpha)}$$

$$= \lim_{t\to 0} \frac{2t\left(\beta^2 e^{t\beta} - \alpha^2 e^{t\alpha}\right) + t^2\left(\beta^3 e^{t\beta} - \alpha^3 e^{t\alpha}\right) - 2\left(\beta e^{t\beta} - \alpha e^{t\alpha}\right) - 2t\left(\beta^2 e^{t\beta} - \alpha^2 e^{t\alpha}\right) + 2\beta e^{t\beta} - 2\alpha e^{t\alpha}}{3t^2(\beta-\alpha)}$$

$$= \lim_{t\to 0} \frac{t^2\left(\beta^3 e^{t\beta} - \alpha^3 e^{t\alpha}\right)}{3t^2(\beta-\alpha)} = \lim_{t\to 0} \frac{\beta^3 e^{t\beta} - \alpha^3 e^{t\alpha}}{3(\beta-\alpha)} = \frac{\beta^3 - \alpha^3}{3(\beta-\alpha)} = \frac{\alpha^2 + \alpha\beta + \beta^2}{3}$$

$$V(X) = \mu_2' - \mu^2 = \frac{\alpha^2 + \alpha\beta + \beta^2}{3} - \left(\frac{\alpha+\beta}{2}\right)^2 = \frac{(\beta-\alpha)^2}{12}$$

5-97.

a) $M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x} dx$

$= \lambda \left(\frac{e^{(t-\lambda)x}}{t-\lambda} \right)\Bigg|_0^\infty$ which is finite only if $t < \lambda$.

$$M_X(t) = \lambda \left(\frac{e^{(t-\lambda)x}}{t-\lambda} \right)\Bigg|_0^\infty = \lambda\left(0 - \frac{1}{t-\lambda}\right) = \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda.$$

b) $E(X) = \mu = \mu_1' = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \left.\frac{\lambda}{(\lambda-t)^2}\right|_{t=0} = \frac{1}{\lambda}$

$$\mu_2' = \left.\frac{d^2 M_X(t)}{dt^2}\right|_{t=0} = \left.\frac{2\lambda}{(\lambda-t)^3}\right|_{t=0} = \frac{2}{\lambda^2}$$

$$V(X) = \mu_2' - \mu^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

5-98.

a) $M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} dx = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{(t-\lambda)x} dx$

$\int_0^\infty x^{r-1} e^{(t-\lambda)x} dx$ is finite only if $t < \lambda$. Besides, we need to use integration by substitution by letting $z = (\lambda - t)x$. Note that the limits of the integration stay the same because $z \to 0$ as $x \to 0$, and $z \to \infty$ as $x \to \infty$. So, we have

$$x = \frac{z}{(\lambda - t)} \quad \text{and} \quad dx = \frac{dz}{(\lambda - t)}$$

$$M_X(t) = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{(t-\lambda)x} dx = \frac{\lambda^r}{\Gamma(r)} \int_0^\infty \left(\frac{z}{\lambda - t}\right)^{r-1} e^{-z} \frac{dz}{\lambda - t} = \frac{\lambda^r}{\Gamma(r)(\lambda - t)^r} \int_0^\infty z^{r-1} e^{-z} dz$$

$$= \frac{\lambda^r}{\Gamma(r)(\lambda - t)^r} \Gamma(r) = \frac{\lambda^r}{(\lambda - t)^r} = \left(\frac{\lambda - t}{\lambda}\right)^{-r} = \left(1 - \frac{t}{\lambda}\right)^{-r}$$

As a result, $M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}$ for $t < \lambda$.

Also note that $\Gamma(r) = \int_0^\infty z^{r-1}e^{-z}dz$ for $r > 0$ by the definition of the gamma function.

b) $E(X) = \mu = \mu_1' = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \left.\left(1-\frac{t}{\lambda}\right)^{-r}\right|_{t=0} = \lambda^r(\lambda - t)^{-r}\Big|_{t=0} = \lambda^r r(\lambda - t)^{-r-1}\Big|_{t=0} = \frac{r}{\lambda}$

$$\mu_2' = \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = r(r+1)\lambda^r(\lambda - t)^{-r-2}\Big|_{t=0} = \frac{r(r+1)}{\lambda^2}$$

$$V(X) = \mu_2' - \mu^2 = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \frac{r}{\lambda^2}$$

5-99.

a) $M_Y(t) = M_{X_1}(t).M_{X_2}(t)...M_{X_r}(t) = \frac{\lambda}{\lambda - t}\frac{\lambda}{\lambda - t}...\frac{\lambda}{\lambda - t} = \left(\frac{\lambda}{\lambda - t}\right)^r$

b) $M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^r = \left(1 - \frac{t}{\lambda}\right)^{-r}$ is the moment-generating function of a gamma distribution. As a result, the random variable Y has a gamma distribution with parameters r and λ.

5-100.

a) $M_Y(t) = M_{X_1}(t)M_{X_2}(t) = \exp\left(\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right) \times \exp\left(\mu_2 t + \frac{\sigma_2^2 t^2}{2}\right)$

$$= \exp\left(\mu_1 t + \frac{\sigma_1^2 t^2}{2} + \mu_2 t + \frac{\sigma_2^2 t^2}{2}\right) = \exp\left((\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)\frac{t^2}{2}\right)$$

b) $M_Y(t) = \exp\left((\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)\frac{t^2}{2}\right)$ is the moment-generating function of a normal distribution with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. As a result, the random variable Y has a normal distribution with parameters $\mu_1 + \mu_2$ and $\sigma_1^2 + \sigma_2^2$.

Supplemental Exercises

5-101. The sum of $\sum_x \sum_y f(x, y) = 1$, $\left(1/4\right) + \left(1/8\right) + \left(1/8\right) + \left(1/4\right) + \left(1/4\right) = 1$

and $f_{XY}(x, y) \geq 0$

a) $P(X < 0.5, Y < 1.5) = f_{XY}(0,1) + f_{XY}(0,0) = 1/8 + 1/4 = 3/8$

b) $P(X \le 1) = f_{XY}(0,0) + f_{XY}(0,1) + f_{XY}(1,0) + f_{XY}(1,1) = 3/4$

c) $P(Y < 1.5) = f_{XY}(0,0) + f_{XY}(0,1) + f_{XY}(1,0) + f_{XY}(1,1) = 3/4$

d) $P(X > 0.5, Y < 1.5) = f_{XY}(1,0) + f_{XY}(1,1) = 3/8$

e) E(X) = 0(3/8) + 1(3/8) + 2(1/4) = 7/8
$V(X) = 0^2(3/8) + 1^2(3/8) + 2^2(1/4) - 7/8^2 = 39/64$
E(Y) = 1(3/8) + 0(3/8) + 2(1/4) = 7/8
$V(Y) = 1^2(3/8) + 0^2(3/8) + 2^2(1/4) - 7/8^2 = 39/64$

f) $f_X(x) = \sum_y f_{XY}(x, y)$ *and* $f_X(0) = 3/8, \quad f_X(1) = 3/8, \quad f_X(2) = 1/4$.

g) $f_{Y|1}(y) = \dfrac{f_{XY}(1, y)}{f_X(1)}$ *and* $f_{Y|1}(0) = \frac{1/8}{3/8} = 1/3, \quad f_{Y|1}(1) = \frac{1/4}{3/8} = 2/3$.

h) $E(Y \mid X = 1) = \sum_{x=1} y f_{Y|X=1}(y) = 0(1/3) + 1(2/3) = 2/3$

i) As is discussed in the chapter, because the range of (X, Y) is not rectangular, X and Y are not independent.

j) E(*XY*) = 1.25, E(*X*) = E(*Y*) = 0.875, V(X) = V(Y) = 0.6094
COV(*X*,*Y*) = E(*XY*) - E(*X*)E(*Y*) = $1.25 - 0.875^2 = 0.4844$

$$\rho_{XY} = \frac{0.4844}{\sqrt{0.6094}\sqrt{0.6094}} = 0.7949$$

5-102. a) $P(X = 2, Y = 4, Z = 14) = \dfrac{20!}{2!4!14!} 0.10^2 0.20^4 0.70^{14} = 0.0631$

b) $P(X = 0) = 0.10^0 0.90^{20} = 0.1216$

c) $E(X) = np_1 = 20(0.10) = 2$

$V(X) = np_1(1 - p_1) = 20(0.10)(0.9) = 1.8$

d) $f_{X|Z=z}(X \mid Z = 19) \dfrac{f_{XZ}(x, z)}{f_Z(z)}$

$$f_{XZ}(xz) = \frac{20!}{x!z!(20 - x - z)!} 0.1^x 0.2^{20-x-z} 0.7^z$$

$$f_Z(z) = \frac{20!}{z!(20 - z)!} 0.3^{20-z} 0.7^z$$

$$f_{X|Z=z}(X \mid Z = 19) \frac{f_{XZ}(x, z)}{f_Z(z)} = \frac{(20 - z)!}{x!(20 - x - z)!} \frac{0.1^x 0.2^{20-x-z}}{0.3^{20-z}} = \frac{(20 - z)!}{x!(20 - x - z)!} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{20-x-z}$$

Therefore, *X* is a binomial random variable with *n*=20-*z* and *p*=1/3. When *z*=19,

$f_{X|19}(0) = \dfrac{2}{3}$ and $f_{X|19}(1) = \dfrac{1}{3}$.

e) $E(X \mid Z = 19) = 0\left(\dfrac{2}{3}\right) + 1\left(\dfrac{1}{3}\right) = \dfrac{1}{3}$

5-103. Let X, Y, and Z denote the number of bolts rated high, moderate, and low. Then, X, Y, and Z have a multinomial distribution.

a) $P(X = 12, Y = 6, Z = 2) = \dfrac{20!}{12!6!2!} 0.6^{12} 0.3^{6} 0.1^{2} = 0.0560$

b) Because X, Y, and Z are multinomial, the marginal distribution of Z is binomial with n = 20 and p = 0.1

c) E(Z) = np = 20(0.1) = 2

d) P(low > 2)=1 – [P(low = 0) + P(low = 1) + P(low = 2)]
= 1 - $[0.9^{20} + 20(0.1^{1})(0.9^{19}) + 190(0.1^{2})(0.9^{18})] = 0.323$

e) $f_{Z|16}(z) = \dfrac{f_{XZ}(16, z)}{f_X(16)}$ and $f_{XZ}(x, z) = \dfrac{20!}{x!\,z!\,(20 - x - z)!} 0.6^{x} 0.3^{(20-x-z)} 0.1^{z}$ for $x + z \le 20$ *and* $0 \le x, 0 \le z$. Then,

$$f_{Z|16}(z) = \frac{\frac{20!}{16!z!(4-z)!} 0.6^{16} 0.3^{(4-z)} 0.1^{z}}{\frac{20!}{16!4!} 0.6^{16} 0.4^{4}} = \frac{4!}{z!(4-z)!}\left(\frac{0.3}{0.4}\right)^{4-z}\left(\frac{0.1}{0.4}\right)^{z}$$

for $0 \le z \le 4$. That is the distribution of Z given X = 16 is binomial with n = 4 and p = 0.25

f) From part (a), E(Z) = 4 (0.25) = 1

g) Because the conditional distribution of Z given X = 16 does not equal the marginal distribution of Z, X and Z are not independent.

5-104. Let X, Y, and Z denote the number of calls answered in two rings or less, three or four rings, and five rings or more, respectively.

a) $P(X = 8, Y = 1, Z = 1) = \dfrac{10!}{8!1!1!} 0.7^{8} 0.25^{1} 0.05^{1} = 0.0649$

b) Let W denote the number of calls answered in four rings or less. Then, W is a binomial random variable with n = 10 and p = 0.95.
Therefore, P(W = 10) = $\binom{10}{10} 0.95^{10} 0.05^{0} = 0.5987$.

c) E(W) = 10(0.95) = 9.5.

d) $f_{Z|8}(z) = \dfrac{f_{XZ}(8, z)}{f_X(8)}$ and $f_{XZ}(x, z) = \dfrac{10!}{x!z!(10 - x - z)!} 0.70^{x} 0.25^{(10-x-z)} 0.05^{z}$ for $x + z \le 10$ *and* $0 \le x, 0 \le z$. Then,

$$f_{Z|8}(z) = \frac{\frac{10!}{8!z!(2-z)!} 0.70^{8} 0.25^{(2-z)} 0.05^{z}}{\frac{10!}{8!2!} 0.70^{8} 0.30^{2}} = \frac{2!}{z!(2-z)!}\left(\frac{0.25}{0.30}\right)^{2-z}\left(\frac{0.05}{0.30}\right)^{z}$$

for $0 \le z \le 2$. That is Z is binomial with n =2 and p = 0.05/0.30 = 1/6.

e) E(Z) given X = 8 is 2(1/6) = 1/3

f) Because the conditional distribution of Z given X = 8 does not equal the marginal distribution of Z, X and Z are not independent.

5-105. $\displaystyle\int_0^3\int_0^2 cx^2 y\,dy\,dx = \int_0^3 cx^2 \left.\frac{y^2}{2}\right|_0^2 dx = 2c \left.\frac{x^3}{3}\right|_0^3 = 18c$. Therefore, c = 1/18.

a) $P(X<1,Y<1)=\int_0^1\int_0^1 \frac{1}{18}x^2 y\,dy\,dx=\int_0^1 \frac{1}{18}x^2 \left.\frac{y^2}{2}\right|_0^1 dx=\left.\frac{1}{36}\frac{x^3}{3}\right|_0^1=\frac{1}{108}$

b) $P(X<2.5)=\int_0^{2.5}\int_0^2 \frac{1}{18}x^2 y\,dy\,dx=\int_0^{2.5} \frac{1}{18}x^2 \left.\frac{y^2}{2}\right|_0^2 dx=\left.\frac{1}{9}\frac{x^3}{3}\right|_0^{2.5}=0.5787$

c) $P(1<Y<2.5)=\int_0^3\int_1^2 \frac{1}{18}x^2 y\,dy\,dx=\int_0^3 \frac{1}{18}x^2 \left.\frac{y^2}{2}\right|_1^2 dx=\left.\frac{1}{12}\frac{x^3}{3}\right|_0^3=\frac{3}{4}$

d)

$$P(X>2,1<Y<1.5)=\int_2^3\int_1^{1.5} \frac{1}{18}x^2 y\,dy\,dx=\int_2^3 \frac{1}{18}x^2 \left.\frac{y^2}{2}\right|_1^{1.5} dx=\left.\frac{5}{144}\frac{x^3}{3}\right|_2^3$$

$$=\frac{95}{432}=0.2199$$

e) $E(X)=\int_0^3\int_0^2 \frac{1}{18}x^3 y\,dy\,dx=\int_0^3 \frac{1}{18}x^3 2\,dx=\left.\frac{1}{9}\frac{x^4}{4}\right|_0^3=\frac{9}{4}$

f) $E(Y)=\int_0^3\int_0^2 \frac{1}{18}x^2 y^2\,dy\,dx=\int_0^3 \frac{1}{18}x^2 \frac{8}{3}dx=\left.\frac{4}{27}\frac{x^3}{3}\right|_0^3=\frac{4}{3}$

g) $f_X(x)=\int_0^2 \frac{1}{18}x^2 y\,dy=\frac{1}{9}x^2 \quad for\ 0<x<3$

h) $f_{Y|X}(y)=\frac{f_{XY}(1,y)}{f_X(1)}=\frac{\frac{1}{18}y}{\frac{1}{9}}=\frac{y}{2}$ for 0 < y < 2.

i) $f_{X|1}(x)=\frac{f_{XY}(x,1)}{f_Y(1)}=\frac{\frac{1}{18}x^2}{f_Y(1)}$ and $f_Y(y)=\int_0^3 \frac{1}{18}x^2 y\,dx=\frac{y}{2} \quad for\ 0<y<2$.

Therefore, $f_{X|1}(x)=\frac{\frac{1}{18}x^2}{1/2}=\frac{1}{9}x^2$ for 0 < x < 3.

5-106. The region $x^2+y^2\le 1$ and 0 < z < 4 is a cylinder of radius 1 (and base area π) and height 4. Therefore, the volume of the cylinder is 4π and $f_{XYZ}(x,y,z)=\frac{1}{4\pi}$ for $x^2+y^2\le 1$ and 0 < z < 4.

a) The region $X^2+Y^2\le 0.5$ is a cylinder of radius $\sqrt{0.5}$ and height 4. Therefore,
$P(X^2+Y^2\le 0.5)=\frac{4(0.5\pi)}{4\pi}=1/2$.

b) The region $X^2+Y^2\le 0.5$ and 0 < z < 2 is a cylinder of radius $\sqrt{0.5}$ and height 2. Therefore,
$P(X^2+Y^2\le 0.5, Z<2)=\frac{2(0.5\pi)}{4\pi}=1/4$

c) $f_{XY|1}(x,y)=\frac{f_{XYZ}(x,y,1)}{f_Z(1)}$ and $f_Z(z)=\iint_{x^2+y^2\le 1}\frac{1}{4\pi}dy\,dx=1/4$

for 0 < z < 4. Then, $f_{XY|1}(x,y)=\frac{1/4\pi}{1/4}=\frac{1}{\pi} \quad for\ x^2+y^2\le 1$.

d) $f_X(x) = \int_0^4 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{4\pi} dydz = \int_0^4 \frac{1}{2\pi}\sqrt{1-x^2}dz = \frac{2}{\pi}\sqrt{1-x^2}$ for -1 < x < 1

e) $f_{Z|0,0}(z) = \dfrac{f_{XYZ}(0,0,z)}{f_{XY}(0,0)}$ and $f_{XY}(x,y) = \int_0^4 \frac{1}{4\pi}dz = 1/\pi$ for $x^2 + y^2 \le 1$. Then,

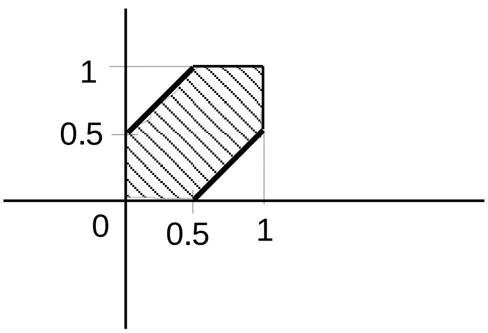
$f_{Z|0,0}(z) = \dfrac{1/4\pi}{1/\pi} = 1/4$ for 0 < z < 4 and $\mu_{Z|0,0} = 2$.

f) $f_{Z|xy}(z) = \dfrac{f_{XYZ}(x,y,z)}{f_{XY}(x,y)} = \dfrac{1/4\pi}{1/\pi} = 1/4$ for 0 < z < 4. Then, E(Z) given X = x and Y = y is

$\int_0^4 \frac{z}{4}dz = 2$.

5-107. $f_{XY}(x,y) = c$ for 0 < x < 1 and 0 < y < 1. Then, $\int_0^1 \int_0^1 cdxdy = 1$ and c = 1.

Because $f_{XY}(x,y)$ is constant, $P(|X - Y| < 0.5)$ is the area of the shaded region below



That is, $P(|X - Y| < 0.5)$ = 3/4.

5-108. a) Let $X_1, X_2, ..., X_6$ denote the lifetimes of the six components, respectively. Because of independence,

$$P(X_1 > 5000, X_2 > 5000, ..., X_6 > 5000) = P(X_1 > 5000)P(X_2 > 5000)...P(X_6 > 5000)$$

If X is exponentially distributed with mean θ, then $\lambda = \frac{1}{\theta}$ and

$P(X > x) = \int_x^\infty \frac{1}{\theta} e^{-t/\theta}dt = -e^{-t/\theta}\Big|_x^\infty = e^{-x/\theta}$. Therefore, the answer is

$e^{-5/8}e^{-0.5}e^{-0.5}e^{-0.25}e^{-0.25}e^{-0.2} = e^{-2.325} = 0.0978$.

b) The probability that at least one component lifetime exceeds 25,000 hours is the same as 1 minus the probability that none of the component lifetimes exceed 25,000 hours. Thus,

1 - P(X_a < 25,000, X_2 < 25,000, …, X_6 < 25,000) = 1 - P(X_1 < 25,000)…P(X_6 < 25,000)
= 1 - (1 - $e^{-25/8}$)(1 - $e^{-2.5}$)(1 - $e^{-2.5}$)(1 - $e^{-1.25}$)(1 - $e^{-1.25}$)(1 - e^{-1}) = 1 - 0.2592 = 0.7408

5-109. Let X, Y, and Z denote the number of problems that result in functional, minor, and no defects, respectively.

a) $P(X=2, Y=5) = P(X=2, Y=5, Z=3) = \frac{10!}{2!5!3!}0.2^2 0.5^5 0.3^3 = 0.085$

b) Z is binomial with n = 10 and p = 0.3

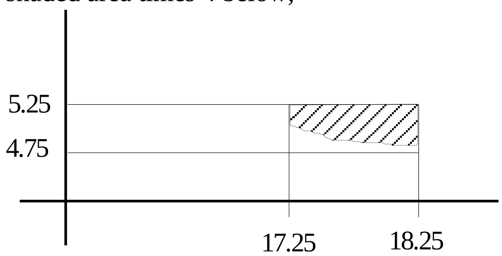
c) E(Z) = 10(0.3) = 3

5-110. a) Let $\overline{X}$ denote the mean weight of the 25 bricks in the sample. Then, E($\overline{X}$) = 3 and $\sigma_{\overline{X}} = \frac{0.25}{\sqrt{25}} = 0.05$. Then, P($\overline{X}$ < 2.95) = P(Z < $\frac{2.95-3}{0.05}$) = P (Z < -1) = 0.159.

b) $P(\overline{X} > x) = P(Z > \frac{x-3}{.05}) = 0.99$. So, $\frac{x-3}{0.05}$ = -2.33 and x=2.8835.

5-111. a)

Because $\int_{17.75}^{18.25}\int_{4.75}^{5.25} cdydx = 0.25c,$ c = 4. The area of a panel is XY and P(XY > 90) is the shaded area times 4 below,



That is, $\int_{17.75}^{18.25}\int_{90/x}^{5.25} 4dydx = 4\int_{17.75}^{18.25} 5.25 - \frac{90}{x}\,dx = 4(5.25x - 90\ln x\Big|_{17.75}^{18.25}) = 0.499$

b) The perimeter of a panel is 2X+2Y and we want P(2X+2Y >46)

$$\int_{17.75}^{18.25}\int_{23-x}^{5.25} 4dydx = 4\int_{17.75}^{18.25} 5.25 - (23-x)dx$$

$$= 4\int_{17.75}^{18.25} (-17.75 + x)dx = 4(-17.75x + \frac{x^2}{2}\Big|_{17.75}^{18.25}) = 0.5$$

5-112. a) Let X denote the weight of a piece of candy and X~N(0.1, 0.01). Each package has 16 candies, then P is the total weight of the package with 16 pieces and E(P) = 16(0.1) = 1.6 ounces and V(P) = $16^2(0.01^2)$ = 0.0256 ounces^2

b) $P(P < 1.6) = P(Z < \frac{1.6-1.6}{0.16}) = P(Z < 0) = 0.5$.

c) Let Y equal the total weight of the package with 17 pieces, E(Y) = 17(0.1) =1.7 ounces and V(Y) = $17^2(0.01^2)$ = 0.0289 ounces^2

$P(Y < 1.6) = P(Z < \frac{1.6-1.7}{\sqrt{0.0289}}) = P(Z < -0.59) = 0.2776$.

5-113. Let $\overline{X}$ denote the average time to locate 10 parts. Then, E($\overline{X}$) =45 and $\sigma_{\overline{X}} = \frac{30}{\sqrt{10}}$

a) $P(\overline{X} > 60) = P(Z > \frac{60-45}{30/\sqrt{10}}) = P(Z > 1.58) = 0.057$

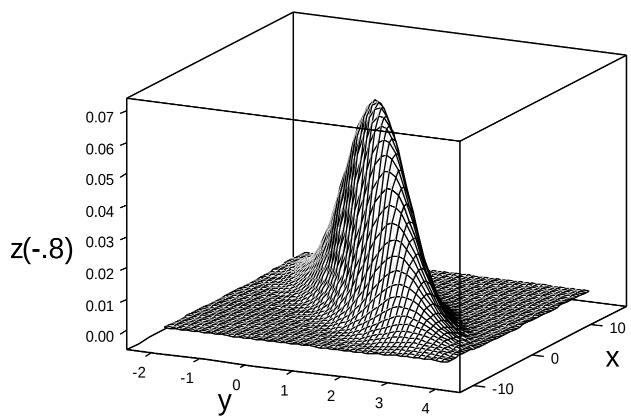
b) Let Y denote the total time to locate 10 parts. Then, Y > 600 if and only if $\overline{X} > 60$. Therefore, the answer is the same as part a.

5-114. a) Let Y denote the weight of an assembly. Then, E(Y) = 4 + 5.5 + 10 + 8 = 27.5 and V(Y)= $0.4^2 + 0.5^2 + 0.2^2 + 0.5^2 = 0.7$.

$$P(Y > 29.5) = P(Z > \tfrac{29.5-27.5}{\sqrt{0.7}}) = P(Z > 2.39) = 0.0084$$

b) Let $\overline{X}$ denote the mean weight of 8 independent assemblies. Then, E($\overline{X}$) = 27.5 and V($\overline{X}$) = 0.7/8 = 0.0875. Also, $P(\overline{X} > 29) = P(Z > \tfrac{29-27.5}{\sqrt{0.0875}}) = P(Z > 5.07) = 0$.

5-115.



5-116.

$$f_{XY}(x,y) = \frac{1}{1.2\pi} e^{\left[\frac{-1}{0.72}\{(x-1)^2 - 1.6(x-1)(y-2) + (y-2)^2\}\right]}$$

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{.36}} e^{\left[\frac{-1}{2(0.36)}\{(x-1)^2 - 1.6(x-1)(y-2) + (y-2)^2\}\right]}$$

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-.8^2}} e^{\left[\frac{-1}{2(1-0.8^2)}\{(x-1)^2 - 2(.8)(x-1)(y-2) + (y-2)^2\}\right]}$$

$E(X) = 1,\ E(Y) = 2\ \ V(X) = 1\ V(Y) = 1$ and $\rho = 0.8$

5-117. Let T denote the total thickness. Then, $T = X_1 + X_2$ and

a) E(T) = 0.5 + 1 = 1.5 mm
$V(T) = V(X_1) + V(X_2) + 2Cov(X_1X_2) = 0.01 + 0.04 + 2(0.014) = 0.078mm^2$
where $Cov(XY) = \rho\sigma_X\sigma_Y = 0.7(0.1)(0.2) = 0.014$

b) $P(T < 1) = P\left(Z < \frac{1-1.5}{\sqrt{0.078}}\right) = P(Z < -1.79) = 0.0367$

c) Let P denote the total thickness. Then, $P = 2X_1 + 3X_2$ and
E(P) = 2(0.5) + 3(1) = 4 mm
$V(P) = 4V(X_1) + 9V(X_2) + 2(2)(3)Cov(X_1X_2)$
$= 4(0.01) + 9(0.04) + 2(2)(3)(0.014) = 0.568mm^2$
where $Cov(XY)=\rho\sigma_X\sigma_Y=0.7(0.1)(0.2)=0.014$

5-118. Let T denote the total thickness. Then, $T = X_1 + X_2 + X_3$ and
a) E(T) = 0.5 + 1 + 1.5 =3 mm
$V(T) = V(X_1) + V(X_2) + V(X_3) + 2Cov(X_1X_2) + 2Cov(X_2X_3) + 2Cov(X_1X_3)$
$= 0.01 + 0.04 + 0.09 + 2(0.014) + 2(0.03) + 2(0.009) = 0.246mm^2$
where $Cov(XY) = \rho\sigma_X\sigma_Y$

b) $P(T < 1.5) = P\left(Z < \frac{1.5-3}{0.246}\right) = P(Z < -6.10) \cong 0$

5-119. Let X and Y denote the percentage returns for security one and two respectively.
If half of the total dollars is invested in each then ½X+ ½Y is the percentage return.
E(½X + ½Y) = 0.05
V(½X + ½Y) = 1/4 V(X) + 1/4V(Y) + 2(1/2)(1/2)Cov(X,Y)
where $Cov(XY)=\rho\sigma_X\sigma_Y=-0.5(2)(4) = -4$

V(½X + ½Y) = 1/4(4) + 1/4(6) - 2 = 3
Also, E(X) = 5 and V(X) = 4.
Therefore, the strategy that splits between the securities has a lower standard deviation of percentage return than investing $2 million in the first security.

5-120. a) The range consists of nonnegative integers with $x + y + z = 4$.
b) Because the samples are selected without replacement, the trials are not independent and the joint distribution is not multinomial.

c)

$$P(X = x \mid Y = 2) = \frac{f_{XY}(x,2)}{f_Y(2)}$$

$$P(Y=2) = \frac{\binom{4}{0}\binom{5}{2}\binom{6}{2}}{\binom{15}{4}} + \frac{\binom{4}{1}\binom{5}{2}\binom{6}{1}}{\binom{15}{4}} + \frac{\binom{4}{2}\binom{5}{2}\binom{6}{0}}{\binom{15}{4}} = 0.1098 + 0.1758 + 0.0440 = 0.3296$$

$$P(X = 0 \text{ and } Y = 2) = \frac{\binom{4}{0}\binom{5}{2}\binom{6}{2}}{\binom{15}{4}} = 0.1098$$

$$P(X = 1 \text{ and } Y = 2) = \frac{\binom{4}{1}\binom{5}{2}\binom{6}{1}}{\binom{15}{4}} = 0.1758$$

$$P(X = 2 \text{ and } Y = 2) = \frac{\binom{4}{1}\binom{5}{2}\binom{6}{1}}{\binom{15}{4}} = 0.0440$$

x	$f_{XY}(x,2)$
0	0.1098/0.3296=0.3331
1	0.1758/0.3296=0.5334
2	0.0440/0.3296=0.1335

d)
P(X = x, Y = y, Z = z) is the number of subsets of size 4 that contain x printers with graphics enhancements, y printers with extra memory, and z printers with both features divided by the number of subsets of size 4.

$$P(X = x, Y = y, Z = z) = \frac{\binom{4}{x}\binom{5}{y}\binom{6}{z}}{\binom{15}{4}} \quad \text{for } x+y+z = 4.$$

$$P(X = 1, Y = 2, Z = 1) = \frac{\binom{4}{1}\binom{5}{2}\binom{6}{1}}{\binom{15}{4}} = 0.1758$$

e) $P(X = 1, Y = 1) = P(X = 1, Y = 1, Z = 2) = \dfrac{\binom{4}{1}\binom{5}{1}\binom{6}{2}}{\binom{15}{4}} = 0.2198$

f) The marginal distribution of X is hypergeometric with N = 15, n = 4, K = 4. Therefore, E(X) = nK/N = 16/15 and V(X) = 4(4/15)(11/15)[11/14] = 0.6146.

g)

$$P(X = 1, Y = 2 \mid Z = 1) = P(X = 1, Y = 2, Z = 1)/P(Z = 1)$$
$$= \left[\frac{\binom{4}{1}\binom{5}{2}\binom{6}{1}}{\binom{15}{4}}\right] \Big/ \left[\frac{\binom{6}{1}\binom{9}{3}}{\binom{15}{4}}\right] = 0.4762$$

h)

$$P(X=2\mid Y=2)=P(X=2,Y=2)/P(Y=2)$$
$$=\left[\frac{\binom{4}{2}\binom{5}{2}\binom{6}{0}}{\binom{15}{4}}\right]\Big/\left[\frac{\binom{5}{2}\binom{10}{2}}{\binom{15}{4}}\right]=0.1334$$

i) Because X + Y + Z = 4, if Y = 0 and Z = 3, then X = 1. Because X must equal 1, $f_{X|YZ}(1)=1$.

5-121. a) Let X, Y, and Z denote the risk of new competitors as no risk, moderate risk, and very high risk. Then, the joint distribution of X, Y, and Z is multinomial with n =12 and $p_1=0.13,\quad p_2=0.72,\quad and\quad p_3=0.15$. X, Y and Z ≥ 0 and x+y+z=12

b) $P(X=1,Y=3,Z=1)$ = 0, not possible since x+y+z ≠ 12

c)
$$P(Z\le 2)=\binom{12}{0}0.15^0 0.85^{12}+\binom{12}{1}0.15^1 0.85^{11}+\binom{12}{2}0.15^2 0.85^{10}$$
$$=0.1422+0.3012+0.2924=0.7358$$

d) $P(Z=2\mid Y=1,X=10)=0$

e)
$$P(X=10)=P(X=10,Y=2,Z=0)+P(X=10,Y=1,Z=1)+P(X=10,Y=0,Z=2)$$
$$=\frac{12!}{10!2!0!}0.13^{10}0.72^2 0.15^0+\frac{12!}{10!1!1!}0.13^{10}0.72^1 0.15^1+\frac{12!}{10!0!2!}0.13^{10}0.72^0 0.15^2$$
$$=4.72x10^{-8}+1.97x10^{-8}+2.04x10^{-9}=6.89x10^{-8}$$
$$P(Z\le 1\mid X=10)=\frac{P(Z=0,Y=2,X=10)}{P(X=10)}+\frac{P(Z=1,Y=1,X=10)}{P(X=10)}$$
$$=\frac{12!}{10!2!0!}0.13^{10}0.72^2 0.15^0\Big/6.89x10^{-8}+\frac{12!}{10!1!1!}0.13^{10}0.72^1 0.15^1\Big/6.89x10^{-8}$$
$$=0.9698$$

f)
$$P(Y\le 1,Z\le 1\mid X=10)=\frac{P(Z=1,Y=1,X=10)}{P(X=10)}$$
$$=\frac{12!}{10!1!1!}0.13^{10}0.72^1 0.15^1\Big/6.89x10^{-8}=0.2852$$

g)
$$E(Z\mid X=10)=(0(4.72x10^{-8})+1(1.97x10^{-8})+2(2.04x10^{-9}))/6.89x10^{-8}$$
$$=0.345$$

5-122.

$$Y=X^{\gamma}\qquad X=\sqrt[\gamma]{Y}$$

Then

$$f_Y(y)=\frac{1}{\sqrt[\gamma]{y}\,\omega\sqrt{2\pi}}e^{\frac{-(\ln(\sqrt[\gamma]{y})-\theta)^2}{2\omega^2}}\left|J\right|,\text{ where } J=\frac{1}{\gamma}y^{\frac{1-\gamma}{\gamma}}$$

Substituting J into $f_Y(y)$, we have $f_Y(y) = \frac{1}{y(\gamma\omega)\sqrt{2\pi}} e^{\frac{-(\ln(y)-\gamma\theta)^2}{2\gamma^2\omega^2}}$

Letting $\omega' = \gamma\omega$, and $\theta' = \gamma\theta$, we have $f_Y(y) = \frac{1}{y\omega'\sqrt{2\pi}} e^{\frac{-(\ln(y)-\theta')^2}{2(\omega')^2}}$

Hence, Y is lognormal with $\theta' = \gamma\theta$, and $\omega' = \gamma\omega$.

5-123.

$$P = \frac{I^2}{R} \qquad I = \sqrt{PR}$$

Then, $f_P(p) = \frac{1}{(0.2)\sqrt{2\pi}} e^{\frac{-(\sqrt{pR}-200)^2}{2(0.2)^2}} \left|J\right|$, where $J = \frac{\sqrt{R}}{2\sqrt{p}}$

5-124.

Because the covariance is zero one may consider a cross section of the beam, say along the x axis. The distribution along the x axis is

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$ and the peak occurs at $x = 0$. Therefore, half the peak is $\frac{1}{2}\frac{1}{\sigma\sqrt{2\pi}}$.

Because the diameter is 1.6 mm, the radius is 0.8 mm at the half peak. Therefore,

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{0.8^2}{2\sigma^2}} = \frac{1}{2}\frac{1}{\sigma\sqrt{2\pi}}$ and $e^{-\frac{0.8^2}{2\sigma^2}} = \frac{1}{2}$

Therefore, upon taking logarithms of both sides $\sigma = 0.679$ mm

5-125.

The moment generating function for a normal distribution is

$$M(t) = \exp\left(\mu t + \frac{\sigma^2}{2}t^2\right)$$

After four derivatives of M(t) we have the following where $\alpha = \mu + \sigma^2 t$ and $\beta = \exp\left(\mu t + \frac{\sigma^2}{2}t\right)$

$$\frac{dM^4(t)}{dt^4} = \alpha^4 \exp(\beta) + 3\sigma^2\alpha^2\exp(\beta) + 2\sigma^2\alpha^2\exp(\beta)$$
$$+2\sigma^4\exp(\beta) + \sigma^2\alpha^2\exp(\beta) + \sigma^4\exp(\beta)$$

And at $t = 0$ we have

$$\frac{dM^4(t)}{dt^4} = \mu^4 + 6\sigma^2\mu^2 + 3\sigma^4$$

For the given values of the normal distribution

E(X) = $\sqrt[4]{E(X^4)} = \sqrt[4]{200^4 + 6(20^2)(200^2) + 3(20^4)} = 202.949$

and this value differs from the mean of 200 by a small amount.

Mind-Expanding Exercises

5-126. By the independence,

$$P(X_1 \in A_1, X_2 \in A_2, ..., X_p \in A_p) = \int_{A_1} \int_{A_2} ... \int_{A_p} f_{X_1}(x_1) f_{X_2}(x_2)... f_{X_p}(x_p) dx_1 dx_2 ... dx_p$$

$$= \left[\int_{A_1} f_{X_1}(x_1) dx_1\right]\left[\int_{A_2} f_{X_2}(x_2) dx_2\right] ... \left[\int_{A_p} f_{X_p}(x_p) dx_p\right]$$

$$= P(X_1 \in A_1) P(X_2 \in A_2) ... P(X_p \in A_p)$$

5-127. $E(Y) = c_1\mu_1 + c_2\mu_2 + ... + c_p\mu_p$. Also,

$$V(Y) = \int \left[c_1x_1 + c_2x_2 + ... + c_px_p - (c_1\mu_1 + c_2\mu_2 + ... + c_p\mu_p)\right]^2 f_{X_1}(x_1) f_{X_2}(x_2)... f_{X_p}(x_p) dx_1 dx_2 ... dx_p$$

$$= \int \left[c_1(x_1 - \mu_1) + ... + c_p(x_p - \mu_p)\right]^2 f_{X_1}(x_1) f_{X_2}(x_2)... f_{X_p}(x_p) dx_1 dx_2 ... dx_p$$

Now, the cross-term

$$\int c_1 c_2 (x_1 - \mu_1)(x_2 - \mu_2) f_{X_1}(x_1) f_{X_2}(x_2)... f_{X_p}(x_p) dx_1 dx_2 ... dx_p$$

$$= c_1 c_2 \left[\int (x_1 - \mu_1) f_{X_1}(x_1) dx_1\right]\left[\int (x_2 - \mu_2) f_{X_2}(x_2) dx_2\right] = 0$$

from the definition of the mean. Therefore, each cross-term in the last integral for V(Y) is zero and

$$V(Y) = \left[\int c_1^2 (x_1 - \mu_1)^2 f_{X_1}(x_1) dx_1\right] ... \left[\int c_p^2 (x_p - \mu_p)^2 f_{X_p}(x_p) dx_p\right]$$

$$= c_1^2 V(X_1) + ... + c_p^2 V(X_p).$$

5-128. $\int_0^a \int_0^b f_{XY}(x,y) dy dx = \int_0^a \int_0^b c dy dx = cab$. Therefore, c = 1/ab. Then, $f_X(x) = \int_0^b c dy = \frac{1}{a}$

for 0 < x < a, and $f_Y(y) = \int_0^a c dx = \frac{1}{b}$ for 0 < y < b. Therefore, $f_{XY}(x,y) = f_X(x) f_Y(y)$ for all x and y and X and Y are independent.

5-129. The marginal density of *X* is

$f_X(x) = \int_0^b g(x)h(u) du = g(x)\int_0^b h(u) du = kg(x)$ where $k = \int_0^b h(u) du$. Also,

$f_Y(y) = lh(y)$ where $l = \int_0^a g(v) dv$. Because $f_{XY}(x,y)$ is a probability density function,

$\int_0^a \int_0^b g(x)h(y) dy dx = \left[\int_0^a g(v) dv\right]\left[\int_0^b h(u) du\right] = 1$. Therefore, *kl* = 1 and

$f_{XY}(x,y) = f_X(x) f_Y(y)$ for all x and y.

5-130. The probability function for X is $P(X=x)=\dfrac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$

The number of ways to select x_j items from N_j is $\binom{Nj}{xj}$.

Therefore, from the multiplication rule the total number of ways to select items to meet the conditions is $\binom{N_1}{x_1}\binom{N_2}{x_2}\ldots\binom{N_k}{x_k}$

The total number of subsets of n items selected from N is $\binom{N}{n}$. Therefore

$$P(X_1=x_1,\ldots X_k=x_k)=\frac{\binom{N_1}{x_1}\binom{N_2}{x_2}\ldots\binom{N_k}{x_k}}{\binom{N}{n}}$$

5-131.

The moment generating function of a normal random variable with mean μ and standard deviation σ is

$$M(t) = \exp\,(\mu t + \frac{\sigma^2}{2}t^2)$$

Let X_i be normally distributed with mean μ_i and standard deviation σ_i and assume the X_i are independent. The moment generating function for $Y = X_1 + X_2 + \ldots + X_p$ is the product of the momebt generating functions for each X_i. That is,

$$M_Y(t) = \exp\left(\sum_{i=1}^{p}\mu_i t + \frac{{\sigma_i}^2}{2}t^2\right) = \exp\left(t\sum_{i=1}^{p}\mu_i + t^2\sum_{i=1}^{p}\frac{{\sigma_i}^2}{2}\right)$$

And this is recognized as the moment generating function for a normal distribution with mean and variance equal to the sum of the μ_i and σ_i^2, respectively.

5-132.

The power series expansion for $\exp(x) = 1 + x + x^2/2! + x^3/3! + \ldots$

Therefore, $E[\exp(tX)] = 1 + tE(X) + t^2E(X^2)/2! + t^3(EX^3)/3! + \ldots$ and this is seen to provide a moment in each term in the series.

The moment generating function for a gamma random variable is

$$M(t) = \left(1-\frac{t}{\tau}\right)^{-r}$$

$$\frac{dM\ \ (t)}{dt} = \left(1-\frac{t}{\tau}\right)^{-r-1}\frac{r}{\tau}$$

and at t = 0 this equals

$$E(X) = \frac{dM\ \ (t)}{dt}\Big|_{t=0} = \frac{r}{\tau}$$

Also

$$\frac{dM^2(t)}{dt^2} = \left(1 - \frac{t}{\tau}\right)^{-r-2} \frac{r(r+1)}{\tau^2}$$

Therefore

$$M(t) = 1 + \frac{r}{\tau}t + \frac{r(r+1)}{\tau^2}\frac{t^2}{2!} + \ \ldots$$

$$E(X^2) = \ \frac{dM^2(t)}{dt^2}\Big|_{t=0} = \frac{r(r+1)}{\tau^2}$$

CHAPTER 6

Section 6-1

6-1. No, usually not. For example, if the sample is {2, 3} the mean is 2.5 which is not an observation in the sample.

6-2. No, it is easy to construct a counter example. For example, {1, 2, 3, 1000}.

6-3. No, usually not. For example, the mean of {1, 4, 4} is 3 which is not even an observation in the sample.

6-4. Yes. For example, {1, 2, 5, 10, 100}, the sample mean = 23.6, sample standard deviation = 42.85

6-5. Yes. For example, {5, 5, 5, 5, 5, 5, 5}, the sample mean = 5, sample standard deviation = 0

6-6. The mean is increased by 10 and the standard deviation is not changed. Try it!

6-7. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{592.035}{8} = 74.0044 \text{ mm}$$

Sample variance:

$$\sum_{i=1}^{8} x_i = 592.035$$

$$\sum_{i=1}^{8} x_i^2 = 43813.18031$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{43813.18031 - \frac{(592.035)^2}{8}}{8-1}$$

$$= \frac{0.0001569}{7} = 0.000022414 \quad (\text{mm})^2$$

Sample standard deviation:

$$s = \sqrt{0.000022414} = 0.00473 \quad \text{mm}$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad \text{where} \quad \sum_{i=1}^{8}(x_i - \bar{x})^2 = 0.0001569$$

Dot Diagram:

```
                   . .  . . . :              .
-------+---------+---------+---------+---------+---------diameter
 73.9920   74.0000   74.0080   74.0160   74.0240   74.0320
```

There appears to be a possible outlier in the data set.

6-8. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{19} x_i}{19} = \frac{272.82}{19} = 14.359 \quad \text{min}$$

Sample variance:

$$\sum_{i=1}^{19} x_i = 272.82$$

$$\sum_{i=1}^{19} x_i^2 = 10333.8964$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{10333.8964 - \frac{(272.82)^2}{19}}{19-1}$$

$$= \frac{6416.49}{18} = 356.47 \quad (\text{min})^2$$

Sample standard deviation:

$$s = \sqrt{356.47} = 18.88 \quad \text{min}$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{19} (x_i - \bar{x})^2 = 6416.49$$

6-9. Sample average:

$$\bar{x} = \frac{84817}{12} = 7068.1 \quad yards$$

Sample variance:

$$\sum_{i=1}^{12} x_i = 84817$$

$$\sum_{i=1}^{19} x_i^2 = 600057949$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{600057949 - \frac{(84817)^2}{12}}{12-1}$$

$$= \frac{564324.92}{11} = 51302.265 \quad (yards)^2$$

Sample standard deviation:

$$s = \sqrt{51302.265} = 226.5 \quad yards$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{12} (x_i - \bar{x})^2 = 564324.92$$

Dot Diagram: (rounding was used to create the dot diagram)

```
       .        .   :   . .  . .     :                                   :
   -+---------+---------+---------+---------+---------+-----C1
  6750      6900      7050      7200      7350      7500
```

6-10. Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{18} x_i}{18} = \frac{2272}{18} = 126.22 \quad \text{kN}$$

Sample variance:

$$\sum_{i=1}^{18} x_i = 2272$$

$$\sum_{i=1}^{18} x_i^2 = 298392$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{298392 - \frac{(2272)^2}{18}}{18-1}$$

$$= \frac{11615.11}{17} = 683.24 \quad (\text{kN})^2$$

Sample standard deviation:

$$s = \sqrt{683.24} = 26.14 \quad \text{kN}$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{18} (x_i - \bar{x})^2 = 11615.11$$

Dot Diagram:

```
         .
     :   ::  .           ::       .          .  : .   .
 +---------+---------+---------+---------+---------+-------yield
90       105       120       135       150       165
```

6-11. Sample average:

$$\bar{x} = \frac{351.8}{8} = 43.975$$

Sample variance:

$$\sum_{i=1}^{8} x_i = 351.8$$

$$\sum_{i=1}^{19} x_i^2 = 16528.403$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{16528.043 - \frac{(351.8)^2}{8}}{8-1}$$

$$= \frac{1057.998}{7} = 151.143$$

Sample standard deviation:

$$s = \sqrt{151.143} = 12.294$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{8}(x_i - \bar{x})^2 = 1057.998$$

Dot Diagram:

```
        .                .  ..    .         ..                         .
    +---------+---------+---------+---------+---------+-------
24.0      32.0      40.0      48.0      56.0      64.0
```

6-12. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{35} x_i}{35} = \frac{28368}{35} = 810.514 \quad watts/m^2$$

Sample variance:

$$\sum_{i=1}^{19} x_i = 28368$$

$$\sum_{i=1}^{19} x_i^2 = 23552500$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{23552500 - \frac{(28368)^2}{35}}{35-1} = \frac{559830.743}{34}$$

$$= 16465.61 \quad (watts/m^2)^2$$

Sample standard deviation:

$$s = \sqrt{16465.61} = 128.32 \quad watts/m^2$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{35} (x_i - \bar{x})^2 = 559830.743$$

Dot Diagram (rounding of the data is used to create the dot diagram)

```
                                    x̄

                             . .
               .     :        .:  ... . . .:  :. . .:. .:: :::
-----+---------+---------+---------+---------+---------+-x
    500       600       700       800       900      1000
```

The sample mean is the point at which the data would balance if it were on a scale.

6-13. $\mu = \frac{6905}{1270} = 5.44$

The value 5.44 is the population mean because the actual physical population of all flight times during the operation is available.

6-14. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{19.56}{9} = 2.173 \quad mm$$

Sample variance:

$$\sum_{i=1}^{9} x_i = 19.56$$

$$\sum_{i=1}^{9} x_i^2 = 45.953$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{45.953 - \frac{(19.56)^2}{9}}{9-1} = \frac{3.443}{8}$$

$$= 0.4303 \quad (mm)^2$$

Sample standard deviation:

$$s = \sqrt{0.4303} = 0.6560 \quad mm$$

Dot Diagram

```
.      .            .      .  .          .  .         ..
-------+---------+---------+---------+---------+---------crack length
    1.40      1.75      2.10      2.45      2.80      3.15
```

6-15. Sample average of exercise group:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3454.68}{12} = 287.89$$

Sample variance of exercise group:

$$\sum_{i=1}^{n} x_i = 3454.68$$

$$\sum_{i=1}^{n} x_i^2 = 1118521.54$$

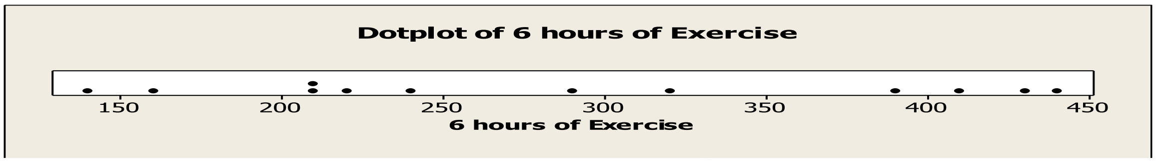
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{1118521.54 - \frac{(3454.68)^2}{12}}{12-1}$$

$$= \frac{123953.71}{11} = 11268.52$$

Sample standard deviation of exercise group:

$$s = \sqrt{11268.52} = 106.15$$

Dot Diagram of exercise group:



Sample average of no exercise group:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2600.08}{8} = 325.010$$

Sample variance of no exercise group:

$$\sum_{i=1}^{n} x_i = 2600.08$$

$$\sum_{i=1}^{n} x_i^2 = 947873.4$$

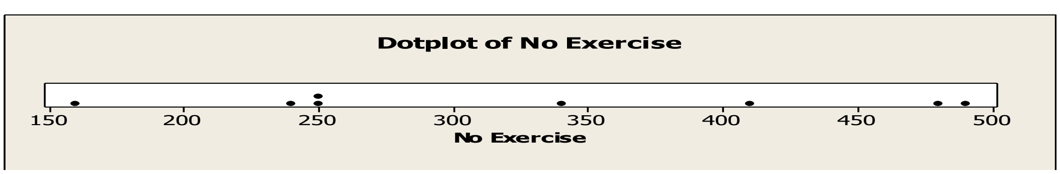
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{947873.4 - \frac{(2600.08)^2}{8}}{8-1}$$

$$= \frac{102821.4}{7} = 14688.77$$

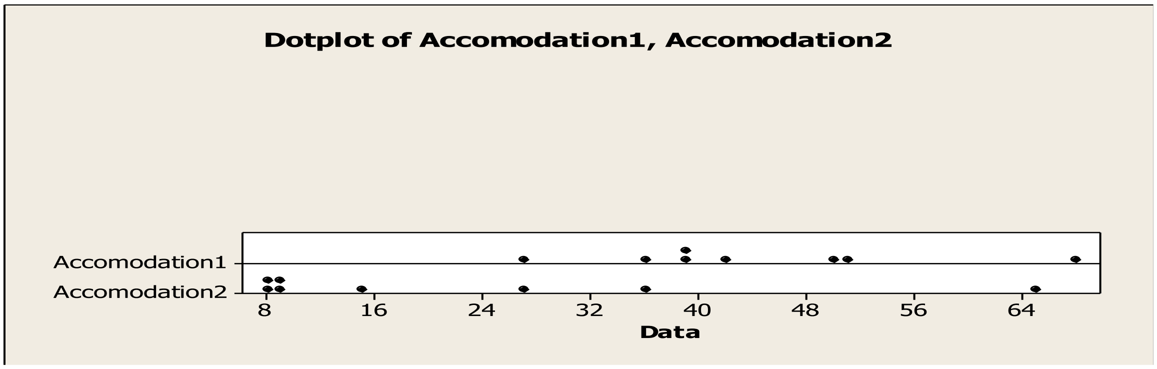
Sample standard deviation of no exercise group:

$$s = \sqrt{14688.77} = 121.20$$

Dot Diagram of no exercise group:



6-16. Dot Diagram of CRT data (Data were rounded for the plot)



The data are distributed at lower values in the second experiment. The lower CRT resolution reduces the visual accommodation.

6-17. $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{57.47}{8} = 7.184$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{412.853 - \frac{(57.47)^2}{8}}{8-1} = \frac{0.00299}{7} = 0.000427$$

$$s = \sqrt{0.000427} = 0.02066$$

Examples: repeatability of the test equipment, time lag between samples, during which the pH of the solution could change, and operator skill in drawing the sample or using the instrument.

6-18. sample mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{748.0}{9} = 83.11$ drag counts

sample variance $s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{62572 - \frac{(748.0)^2}{9}}{9-1}$

$$= \frac{404.89}{8} = 50.61 \text{ drag counts}^2$$

sample standard deviation $s = \sqrt{50.61} = 7.11$ drag counts

Dot Diagram

```
  .         . . . . . . .                             .
---+---------+---------+---------+---------+---------+---drag counts
75.0      80.0      85.0      90.0      95.0     100.0
```

6-19. a) $\bar{x} = 65.86 \ ^\circ F$
$s = 12.16 \ ^\circ F$

Dot Diagram

```
                                      :  :
  .        .    .   .  ..   .: .: .  .:..: .. :: .... ..
-+---------+---------+---------+---------+---------+-----temp
30        40        50        60        70        80
```

b) Removing the smallest observation (31), the sample mean and standard deviation become

$\bar{x} = 66.86\ °F$

$s = 10.74\ °F$

6-20. Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{159.85}{30} = 5.32833$$

Sample variance:

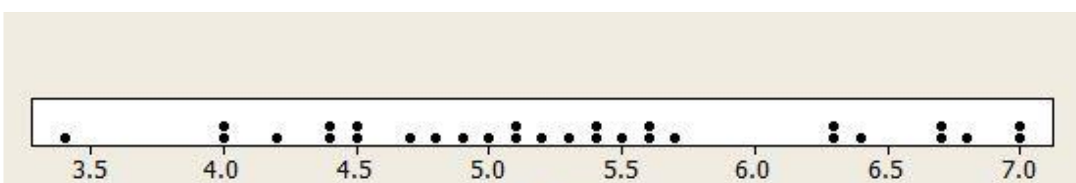
$$\sum_{i=1}^{30} x_i = 159.85 \qquad \sum_{i=1}^{30} x_i^2 = 879.3213$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{879.3213 - \frac{(159.85)^2}{30}}{30-1} = \frac{27.58722}{29} = 0.951283$$

Sample standard deviation:

$$s = \sqrt{0.951283} = 0.975338$$

Dot Diagram:



6-21. Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{188.82}{40} = 4.7205$$

Sample variance:

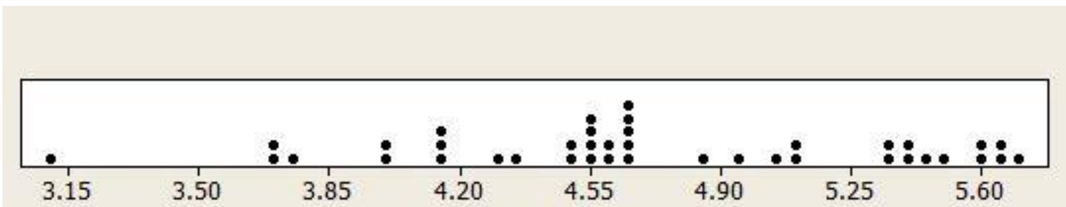
$$\sum_{i=1}^{40} x_i = 188.82 \qquad \sum_{i=1}^{40} x_i^2 = 907.157$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{907.157 - \frac{(188.82)^2}{40}}{40-1} = \frac{15.8322}{39} = 0.405954$$

Sample standard deviation:

$$s = \sqrt{0.405954} = 0.637145$$

Dot Diagram:



6-22. a) All 52 clouds:

Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{15770.9}{52} = 303.2865$$

Sample variance:

$$\sum_{i=1}^{52} x_i = 15770.9 \qquad \sum_{i=1}^{52} x_i^2 = 18309564$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{18309564 - \frac{(15770.9)^2}{52}}{52-1} = \frac{13526462}{51} = 265224.8$$

Sample standard deviation:

$$s = \sqrt{265224.8} = 514.9998$$

Range:

Maximum= 2745.6, Minimum= 1, Range= 2745.6 - 1 = 2744.6

b) Unseeded Clouds:

Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4279.3}{26} = 164.588$$

Sample variance:

$$\sum_{i=1}^{26} x_i = 4279.3 \qquad \sum_{i=1}^{26} x_i^2 = 2642355$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{2642355 - \frac{(4279.3)^2}{26}}{26-1} = \frac{1938032}{25} = 77521.26$$

Sample standard deviation:

$$s = \sqrt{77521.26} = 278.4264$$

Range:

Maximum= 1202.6, Minimum= 1, Range= 1202.6-1 = 1201.6

c) Seeded Clouds:

Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{11491.6}{26} = 441.985$$

Sample variance:

$$\sum_{i=1}^{26} x_i = 11491.6 \qquad \sum_{i=1}^{26} x_i^2 = 15667208.96$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{15667208.96 - \frac{(11491.6)^2}{26}}{26-1} = \frac{10588099}{25} = 423524$$

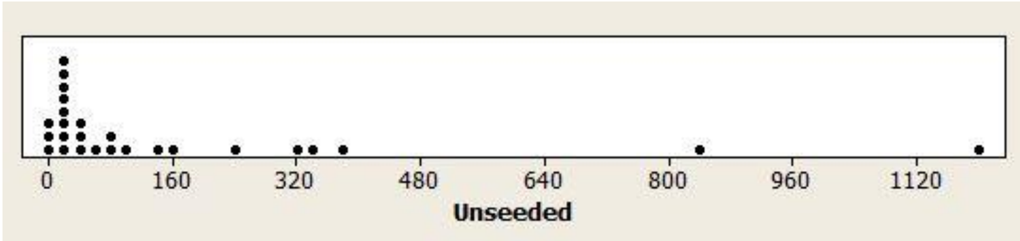
Sample standard deviation:

$$s = \sqrt{423524} = 650.787$$

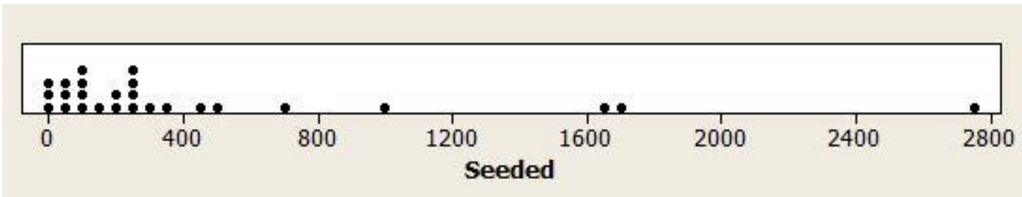
Range:

Maximum= 2745.6, Minimum= 4.1, Range= 2745.6-4.1 = 2741.5

6-23. Dot Diagram for Unseeded clouds:



Dot Diagram for Seeded clouds:



The sample mean for unseeded clouds is 164.588 and the data is not centered about the mean. Two large observations increase the mean. The sample mean for seeded clouds is 441.985 and the data is not centered about the mean. The average rainfall when clouds are seeded is higher when compared to the rainfall when clouds are not seeded. The amount of rainfall of seeded clouds varies widely when compared to amount of rainfall for unseeded clouds.

6-24. Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3737.98}{71} = 52.6476$$

Sample variance:

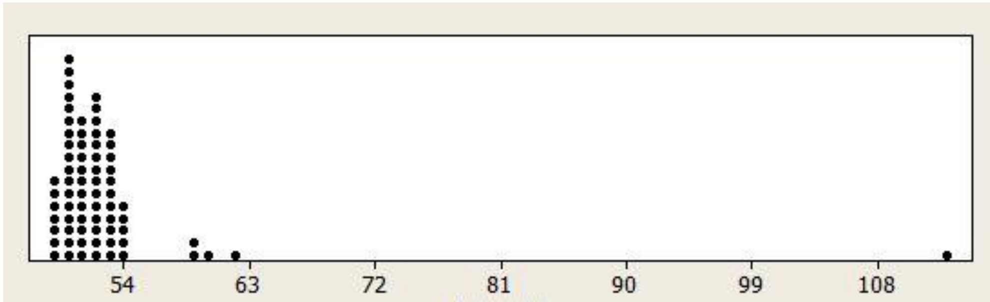
$$\sum_{i=1}^{71} x_i = 3737.98 \qquad \sum_{i=1}^{71} x_i^2 = 200899$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{200899 - \frac{(3737.98)^2}{71}}{71-1} = \frac{4102.98}{70} = 58.614$$

Sample standard deviation:

$$s = \sqrt{58.614} = 7.65598$$

Dot Diagram:



There appears to be an outlier in the data.

Section 6-2

6-25. a) N = 30

```
Leaf Unit = 0.10


 1    3  4
 2    3  9
 6    4  0134
 12   4  556899
(7)   5  0012444
 11   5  567
 8    6  234
 5    6  678
 2    7  00

Variable   N  N*   Mean  SE Mean  StDev  Minimum     Q1  Median     Q3
CondNum   30  0  5.328    0.178  0.975    3.420  4.538   5.220  6.277
Maximum
7.000
```

b) One particular bridge has poor rating of 3.4

c) Calculating the mean after removing the bridge mentioned above:

```
Variable   N   Mean  SE Mean  StDev  Minimum     Q1  Median     Q3
CondNum   29  5.394    0.171  0.922    3.970  4.600   5.260  6.295
Maximum
7.000
```

There is a little difference in between the two means.

6-26. a) N = 40

```
Leaf Unit = 0.10


 1    3  1
 1    3
 1    3
 4    3  777
 6    3  99
 9    4  111
 11   4  33
 17   4  555555
```

```
(7)   4   6666666
 16   4   89
 14   5   011
 11   5   3333
 7    5   445
 4    5   6667
```

Descriptive Statistics: pH

```
Variable    N   N*    Mean   SE Mean   StDev   Minimum      Q1   Median      Q3
pH         40    0   4.720     0.101   0.637     3.100   4.325    4.635   5.370
Maximum
5.700
```

b) The number of observations below 5.3 is 17 + 7 + 5 = 29 observations
Therefore, the percentage of observations which are considered acid rain is 72.5%

6-27. A back-to-back steam-and-leaf display is useful when two data sets are to be compared. Therefore, for this problem the comparison of the seeded versus unseeded clouds can be performed more easily with the back-to-back stem and leaf diagram than a dot diagram.

6-28. The median will be equal to the mean when the sample is symmetric about the mean value.

6-29. The median will equal the mode when the sample is symmetric with a single mode. The symmetry implies the mode is at the median of the sample.

6-30.

```
Stem-and-leaf of C1  N  = 82
Leaf Unit = 0.10


 1    83   4
 3    84   33
 4    85   3
 7    86   777
 13   87   456789
 24   88   23334556679
 34   89   0233678899
(13)  90   0111344456789
 35   91   0001112256688
 22   92   22236777
 14   93   023347
 8    94   2247
 4    95
 4    96   15
 2    97
 2    98   8
 1    99
 1    100  3
Q1           Median        Q3
88.575       90.400        92.200
```

6-31. Stem-and-leaf display for cycles to failure: `unit = 100     1|2  represents 1200`

```
  1    0T|3
  1    0F|
  5    0S|7777
 10    0o|88899
 22    1*|000000011111
```

```
 33    1T|22222223333
(15)   1F|444445555555555
 22    1S|66667777777
 11    1o|888899
  5    2*|011
  2    2T|22
```

Median = 1436.5, Q_1 = 1097.8, and Q_3 = 1735.0

No, only 5 out of 70 coupons survived beyond 2000 cycles.

6-32.
```
Stem-and-leaf display of percentage of cotton   N  = 64
Leaf Unit = 0.10        32|1 represents 32.1%

  1   32|1
  6   32|56789
  9   33|114
 17   33|56666688
 24   34|0111223
(14)  34|55666667777779
 26   35|001112344
 17   35|56789
 12   36|234
  9   36|6888
  5   37|13
  3   37|689
```

Median = 34.7, Q_1 = 33.800, and Q_3 = 35.575

6-33. Stem-and-leaf display for yield: `unit = 1    1|2  represents 12`

```
  1    7o|8
  1    8*|
  7    8T|223333
 21    8F|44444444555555
 38    8S|66666666667777777
(11)   8o|88888999999
 41    9*|00000000001111
 27    9T|22233333
 19    9F|444444445555
  7    9S|666677
  1    9o|8
```

Median = 89.250, Q_1 = 86.100, and Q_3 = 93.125

6-34. The data in the 42nd is 90.4 which is median.

The mode is the most frequently occurring data value. There are several data values that occur 3 times. These are: 86.7, 88.3, 90.1, 90.4, 91, 91.1, 91.2, 92.2, and 92.7, so this data set has a multimodal distribution.

Sample mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{7514.3}{83} = 90.534$

6-35. Sample median is at $\frac{(70+1)}{2}$ = 35.5th observation, the median is 1436.5.

Modes are 1102, 1315, and 1750 which are the most frequent data.

Sample mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{98259}{70} = 1403.7$

6-36. Sample median is at $\frac{(64+1)}{2}$ = 32.5th

The 32nd is 34.7 and the 33rd is 34.7, so the median is 34.7.

Mode is 34.7 which is the most frequent data.

Sample mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2227.1}{64} = 34.798$

6-37. Do not use the total as an observation. There are 23 observations.

```
Stem-and-leaf of Billion of kilowatt hours  N  = 23
Leaf Unit = 100


(18)  0  000000000000000111
 5    0  23
 3    0  5
 2    0
 2    0  9
 1    1
 1    1
 1    1
 1    1  6
```

Sample median is at 12th = 38.43.

Sample mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4398.8}{23} = 191.0$

Sample variance: s^2 = 150673.8

Sample standard deviation: s = 388.2

6-38. Sample mean: $\bar{x} = 65.811$ inches, standard deviation $s = 2.106$ inches, and sample median: $\tilde{x} = 66.000$ inches

Stem-and-leaf display of female engineering student heights N = 37

```
Leaf Unit = 0.10  61|0 represents 61.0 inches

   1    61|0
   3    62|00
   5    63|00
   9    64|0000
  17    65|00000000
 (4)    66|0000
  16    67|00000000
   8    68|00000
   3    69|00
   1    70|0
```

6-39. Stem-and-leaf display. `Strength: unit = 1.0 1|2 represents 12`

```
   1  532|9
   1  533|
   2  534|2
   4  535|47
   5  536|6
   9  537|5678
  20  538|12345778888
  26  539|016999
  37  540|11166677889
  46  541|123666688
(13)  542|0011222357899
  41  543|011112556
  33  544|00012455678
  22  545|2334457899
  13  546|23569
   8  547|357
   5  548|11257
```

$$\frac{i-0.5}{100}\times 100 = 95 \Rightarrow i = 95.5 \Rightarrow 95^{th} \ \ percentile \ \ is\, 5479$$

6-40. Stem-and-leaf of concentration, `N = 60, Leaf Unit = 1.0, 2|9 represents 29`
Note: Minitab has dropped the value to the right of the decimal to make this display.

```
   1     2|9
   2     3|1
   3     3|9
   8     4|22223
  12     4|5689
  20     5|01223444
(13)     5|5666777899999
  27     6|11244
  22     6|556677789
  13     7|022333
   7     7|6777
   3     8|01
   1     8|9
```

The data have a symmetrical bell-shaped distribution, and therefore may be normally distributed.

Sample Mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3592.0}{60} = 59.87$

Sample Standard Deviation

$$\sum_{i=1}^{60} x_i = 3592.0 \quad \text{and} \quad \sum_{i=1}^{60} x_i^2 = 224257$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{224257 - \frac{(3592.0)^2}{60}}{60-1} = \frac{9215.93}{59}$$

$$= 156.20$$

and

$$s = \sqrt{156.20} = 12.50$$

Sample Median $\tilde{x} = 59.45$

```
Variable            N     Median
concentration      60      59.45
```

$$\frac{i-0.5}{60} \times 100 = 90 \Rightarrow i = 54.5 \Rightarrow 90^{th} \ \ percentile\ \ is\ 76.85$$

6-41. Stem-and-leaf display. Yard: `unit = 1.0`
Note: Minitab has dropped the value to the right of the decimal to make this display.

1	22 | 6
5	23 | 2334
8	23 | 677
16	24 | 00112444
20	24 | 5578
33	25 | 0111122334444
46	25 | 5555556677899
(15)	26 | 000011123334444
39	26 | 56677888
31	27 | 0000112222233333444
12	27 | 66788999
4	28 | 003
1	28 | 5

Sample Mean $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{\sum_{i=1}^{100} x_i}{100} = \dfrac{26030.2}{100} = 260.3 \quad \textit{yards}$

Sample Standard Deviation

$$\sum_{i=1}^{100} x_i = 26030.2 \qquad \text{and} \qquad \sum_{i=1}^{100} x_i^2 = 6793512$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{6793512 - \frac{((26030.2)^2}{100}}{100-1} = \frac{17798.42}{99}$$

$$= 179.782 \quad \textit{yards}^2$$

and

$$s = \sqrt{179.782} = 13.41 \quad \textit{yards}$$

Sample Median

```
Variable            N     Median
yards             100     260.85
```

$$\frac{i - 0.5}{100} \times 100 = 90 \Rightarrow i = 90.5 \Rightarrow 90^{th} \ \ \textit{percentile is}\ 277.2$$

6-42. Stem-and-leaf of speed (in megahertz) `N = 120`

```
Leaf Unit = 1.0   63|4 represents 634 megahertz

    2   63|47
    7   64|24899
   16   65|223566899
   35   66|0000001233455788899
   48   67|0022455567899
  (17)  68|00001111233333458
   55   69|0000112345555677889
   36   70|011223444556
   24   71|0057889
   17   72|000012234447
    5   73|59
    3   74|68
    1   75|
    1   76|3
```

35/120 = 29% exceed 700 megahertz.

Sample Mean $\bar{x} = \dfrac{\sum_{i=1}^{120} x_i}{120} = \dfrac{82413}{120} = 686.78 \quad mhz$

Sample Standard Deviation

$$\sum_{i=1}^{120} x_i = 82413 \qquad \text{and} \qquad \sum_{i=1}^{120} x_i^2 = 56677591$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{56677591 - \frac{(82413)^2}{120}}{120-1} = \frac{78402.925}{119}$$

$$= 658.85 \quad mhz^2$$

and

$$s = \sqrt{658.85} = 25.67 \quad mhz$$

Sample Median $\tilde{x} = 683.0\, mhz$

```
Variable            N     Median
speed             120     683.00
```

6-43. Stem-and-leaf display. Rating: `unit = 0.10  1|2  represents 1.2`

```
  1    83|0
  2    84|0
  5    85|000
  7    86|00
  9    87|00
 12    88|000
 18    89|000000
 (7)   90|0000000
 15    91|0000000
  8    92|0000
  4    93|0
  3    94|0
  2    95|00
```

Sample Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{40} x_i}{40} = \frac{3578}{40} = 89.45$$

Sample Standard Deviation

$$\sum_{i=1}^{40} x_i = 3578 \qquad \text{and} \qquad \sum_{i=1}^{40} x_i^2 = 320366$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{320366 - \frac{(3578)^2}{40}}{40-1} = \frac{313.9}{39}$$

$$= 8.05$$

and

$$s = \sqrt{8.05} = 2.8$$

Sample Median

```
Variable              N        Median
rating               40        90.000
```

22/40 = 55% of the taste testers considered this particular Pinot Noir truly exceptional.

6-44. Stem-and-leaf diagram of $NbOCl_3$ `N = 27`

```
Leaf Unit = 100          0|4   represents 40 gram-mole/liter x 10^-3

  6    0|444444
  7    0|5
 (9)   1|001122233
 11    1|5679
  7    2|
  7    2|5677
  3    3|124
```

Sample mean $\bar{x} = \dfrac{\sum_{i=1}^{27} x_i}{27} = \dfrac{41553}{27} = 1539 \text{ gram} - \text{mole/liter } \text{x}10^{-3}$

Sample Standard Deviation

$$\sum_{i=1}^{27} x_i = 41553 \qquad \text{and} \qquad \sum_{i=1}^{27} x_i^2 = 87792869$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{87792869 - \frac{(41553)^2}{27}}{27-1} = \frac{23842802}{26} = 917030.85$$

and $s = \sqrt{917030.85} = 957.62 \text{ gram - mole/liter } \text{x}\,10^{-3}$

Sample Median $\tilde{x} = 1256 \text{ gram} - \text{mole/liter } \text{x}10^{-3}$

Variable	N	Median
$NbOCl_3$	40	1256

6-45. Stem-and-leaf display. Height: `unit = 0.10    1|2 represents 1.2`

```
  Female Students|| Male Students
        0|61   1  |
       00|62   3  |
       00|63   5  |
     0000|64   9  |
 00000000|65  17  |  2   65|00
     0000|66  (4) |  3   66|0
 00000000|67  16  |  7   67|0000
    00000|68   8  | 17   68|0000000000
       00|69   3  |(15)  69|000000000000000
        0|70   1  | 18   70|0000000
                  | 11    71|00000
                  |  6    72|00
                  |  4   73|00
                  |  2  74|0
                  |  1   75|0
```

The male engineering students are taller than the female engineering students. Also there is a slightly wider range in the heights of the male students.

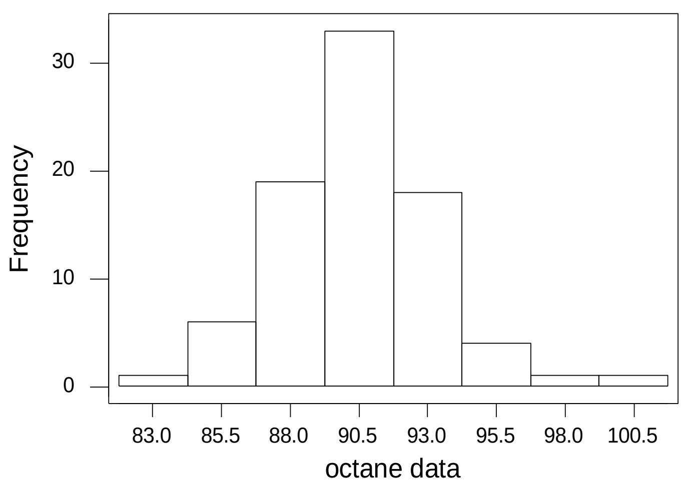
Section 6-3

6-46.

Solution uses the $n = 83$ observations from the data set.

Frequency Tabulation for Exercise 6-22.Octane Data

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		81.75		0	.0000	0	.0000
1	81.75	84.25	83.0	1	.0120	1	.0120
2	84.25	86.75	85.5	6	.0723	7	.0843
3	86.75	89.25	88.0	19	.2289	26	.3133
4	89.25	91.75	90.5	33	.3976	59	.7108
5	91.75	94.25	93.0	18	.2169	77	.9277
6	94.25	96.75	95.5	4	.0482	81	.9759
7	96.75	99.25	98.0	1	.0120	82	.9880
8	99.25	101.75	100.5	1	.0120	83	1.0000
above	101.75			0	.0000	83	1.0000



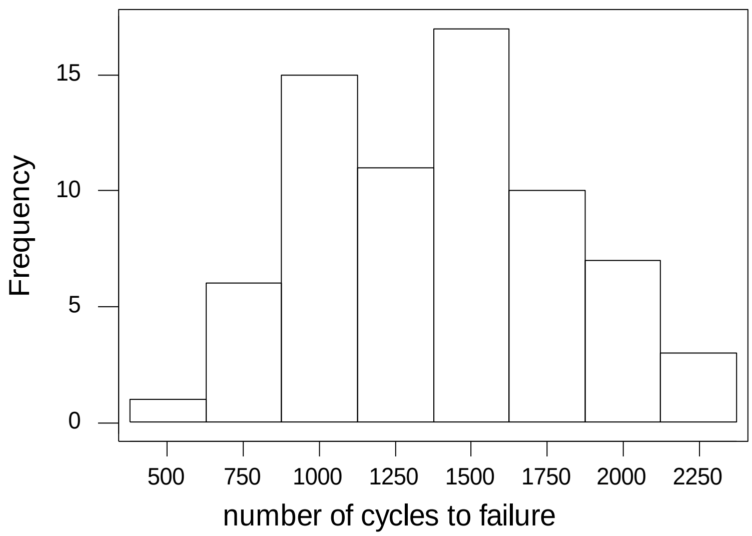
Mean = 90.534 Standard Deviation = 2.888 Median = 90.400

6-47.

Frequency Tabulation for Exercise 6-23.Cycles

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		.000		0	.0000	0	.0000
1	.000	266.667	133.333	0	.0000	0	.0000
2	266.667	533.333	400.000	1	.0143	1	.0143
3	533.333	800.000	666.667	4	.0571	5	.0714
4	800.000	1066.667	933.333	11	.1571	16	.2286
5	1066.667	1333.333	1200.000	17	.2429	33	.4714
6	1333.333	1600.000	1466.667	15	.2143	48	.6857
7	1600.000	1866.667	1733.333	12	.1714	60	.8571
8	1866.667	2133.333	2000.000	8	.1143	68	.9714
9	2133.333	2400.000	2266.667	2	.0286	70	1.0000
above	2400.000			0	.0000	70	1.0000

Mean = 1403.66 Standard Deviation = 402.385 Median = 1436.5

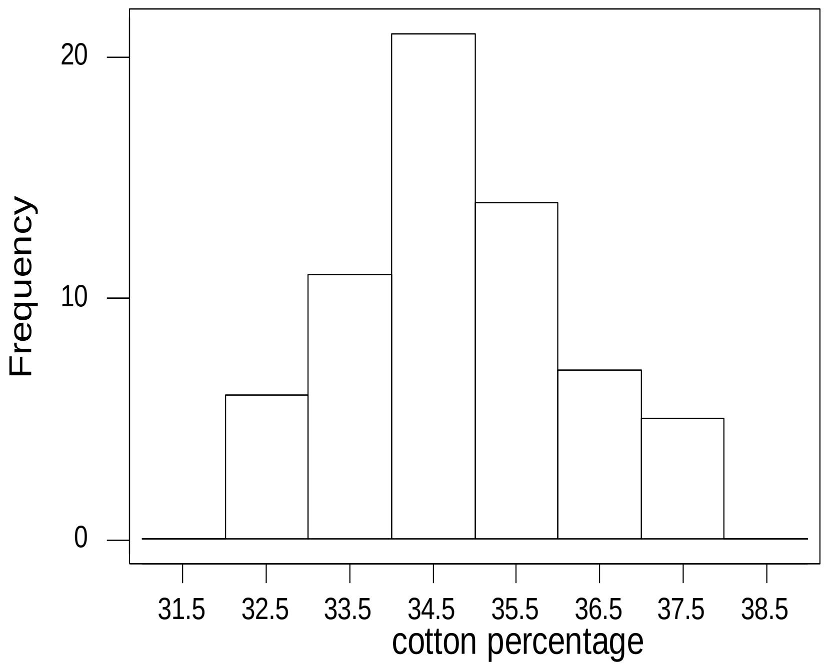


6-48.

Frequency Tabulation for Exercise 6-24.Cotton content

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		31.0		0	.0000	0	.0000
1	31.0	32.0	31.5	0	.0000	0	.0000
2	32.0	33.0	32.5	6	.0938	6	.0938
3	33.0	34.0	33.5	11	.1719	17	.2656
4	34.0	35.0	34.5	21	.3281	38	.5938
5	35.0	36.0	35.5	14	.2188	52	.8125
6	36.0	37.0	36.5	7	.1094	59	.9219
7	37.0	38.0	37.5	5	.0781	64	1.0000
8	38.0	39.0	38.5	0	.0000	64	1.0000
above	39.0			0	.0000	64	1.0000

Mean = 34.798 Standard Deviation = 1.364 Median = 34.700

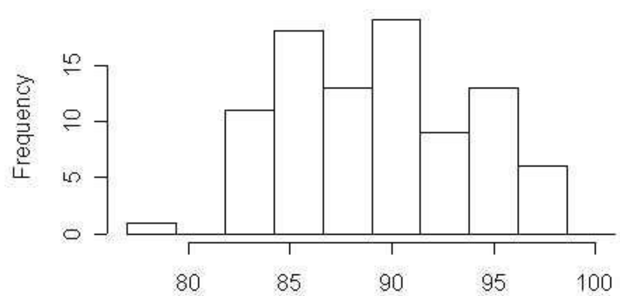


6-49.

Frequency Tabulation for Exercise 6-25.Yield

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		77.000		0	.0000	0	.0000
1	77.000	79.400	78.200	1	.0111	1	.0111
2	79.400	81.800	80.600	0	.0000	1	.0111
3	81.800	84.200	83.000	11	.1222	12	.1333
4	84.200	86.600	85.400	18	.2000	30	.3333
5	86.600	89.000	87.800	13	.1444	43	.4778
6	89.000	91.400	90.200	19	.2111	62	.6889
7	91.400	93.800	92.600	9	.1000	71	.7889
8	93.800	96.200	95.000	13	.1444	84	.9333
9	96.200	98.600	97.400	6	.0667	90	1.0000
10	98.600	101.000	99.800	0	.0000	90	1.0000
above	101.000			0	.0000	90	1.0000

Mean = 89.3756 Standard Deviation = 4.31591 Median = 89.25



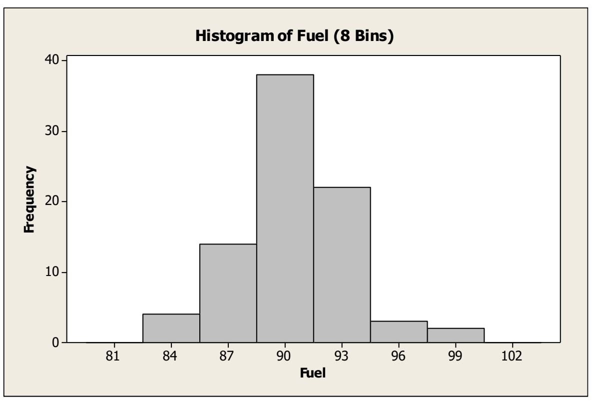
6-50. Solutions uses the n = 83 observations from the data set.

Frequency Tabulation for Exercise 6-22.Octane Data

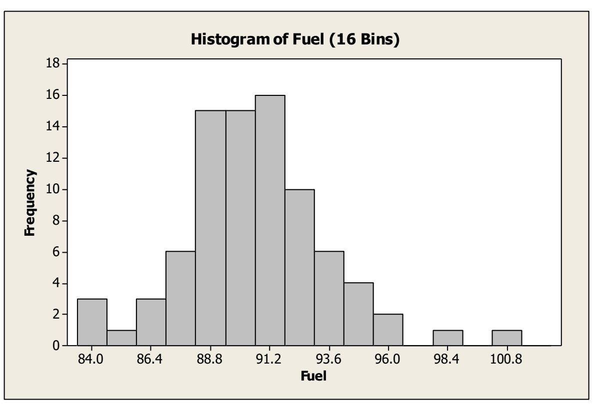
Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		83.000		0	.0000	0	.0000
1	83.000	84.125	83.5625	1	.0120	1	.0120
2	84.125	85.250	84.6875	2	.0241	3	.0361
3	85.250	86.375	85.8125	1	.0120	4	.0482
4	86.375	87.500	86.9375	5	.0602	9	.1084
5	87.500	88.625	88.0625	13	.1566	22	.2651
6	88.625	89.750	89.1875	8	.0964	30	.3614
7	89.750	90.875	90.3125	16	.1928	46	.5542
8	90.875	92.000	91.4375	15	.1807	61	.7349
9	92.000	93.125	92.5625	9	.1084	70	.8434
10	93.125	94.250	93.6875	7	.0843	77	.9277
11	94.250	95.375	94.8125	2	.0241	79	.9518
12	95.375	96.500	95.9375	2	.0241	81	.9759
13	96.500	97.625	97.0625	0	.0000	81	.9759
14	97.625	98.750	98.1875	0	.0000	81	.9759
15	98.750	99.875	99.3125	1	.0120	82	.9880
16	99.875	101.000	100.4375	1	.0120	83	1.0000
above	101.000			0	.0000	83	1.0000

Mean = 90.534 Standard Deviation = 2.888 Median = 90.400

Histogram 8 bins:



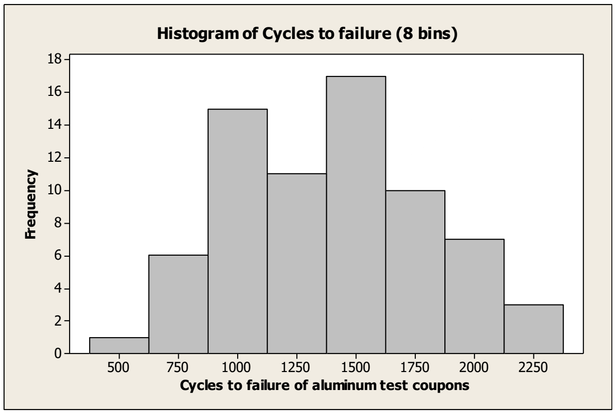
Histogram 16 Bins:



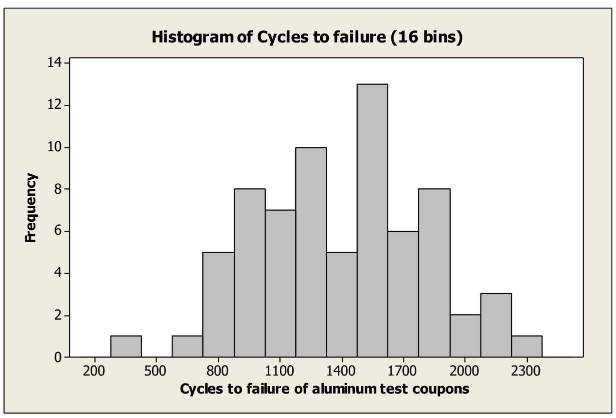
Yes, both of them give the similar information.

6-51.

Histogram 8 bins:



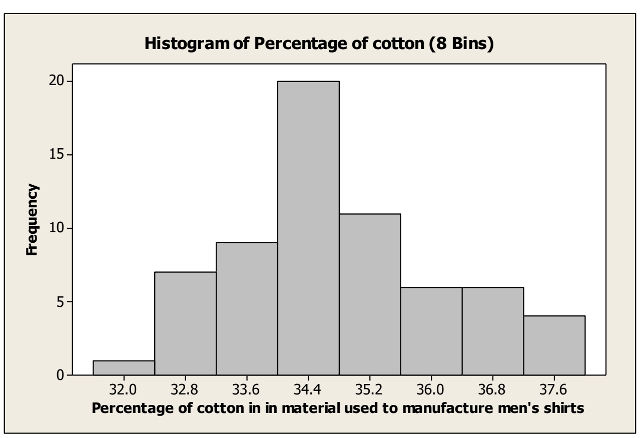
Histogram 16 bins:



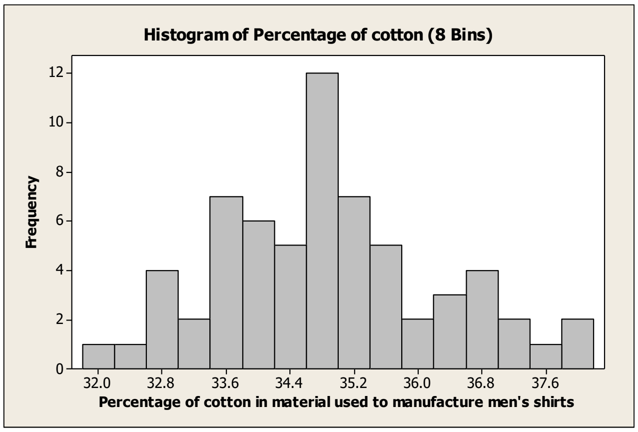
Yes, both of them give the same similar information

6-52.

Histogram 8 bins:



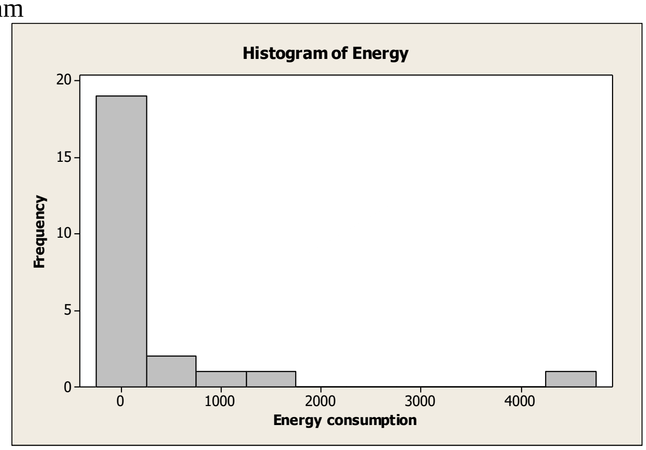
Histogram 16 Bins:



Yes, both of them give similar information.

6-53.

Histogram



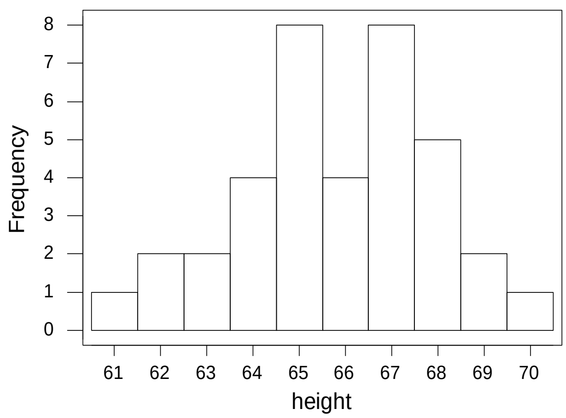
The data are skewed.

6-54.

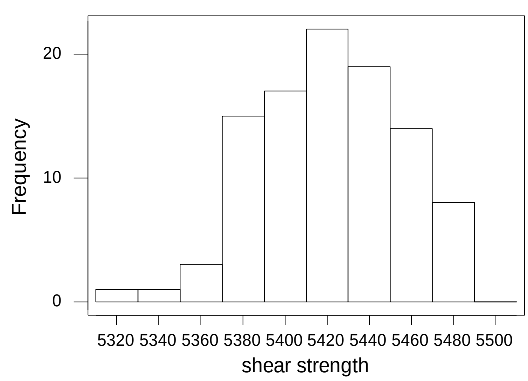
Frequency Tabulation for Problem 6-30. Height Data

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		60.500		0	.0000	0	.0000
1	60.500	61.500	61.000	1	.0270	1	.0270
2	61.500	62.500	62.000	2	.0541	3	.0811
3	62.500	63.500	63.000	2	.0541	5	.1351
4	63.500	64.500	64.000	4	.1081	9	.2432
5	64.500	65.500	65.000	8	.2162	17	.4595
6	65.500	66.500	66.000	4	.1081	21	.5676
7	66.500	67.500	67.000	8	.2162	29	.7838
8	67.500	68.500	68.000	5	.1351	34	.9189
9	68.500	69.500	69.000	2	.0541	36	.9730
10	69.500	70.500	70.000	1	.0270	37	1.0000
above	70.500			0	.0000	37	1.0000

Mean = 65.811   Standard Deviation = 2.106   Median = 66.0



6-55. The histogram for the spot weld shear strength data shows that the data appear to be normally distributed (the same shape that appears in the stem-leaf-diagram).

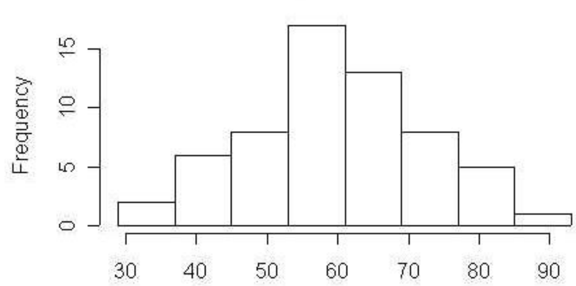


6-56.

Frequency Tabulation for exercise 6-32. Concentration data

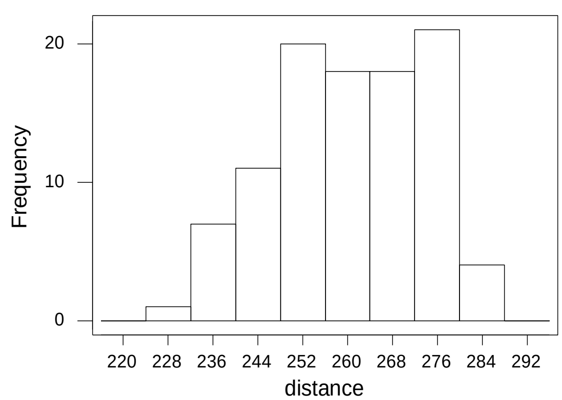
Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		29.000		0	.0000	0	.0000
1	29.0000	37.000	33.000	2	.0333	2	.0333
2	37.0000	45.000	41.000	6	.1000	8	.1333
3	45.0000	53.000	49.000	8	.1333	16	.2667
4	53.0000	61.000	57.000	17	.2833	33	.5500
5	61.0000	69.000	65.000	13	.2167	46	.7667
6	69.0000	77.000	73.000	8	.1333	54	.9000
7	77.0000	85.000	81.000	5	.0833	59	.9833
8	85.0000	93.000	89.000	1	.0167	60	1.0000
above	93.0000			0	.0800	60	1.0000

Mean = 59.87 Standard Deviation = 12.50 Median = 59.45

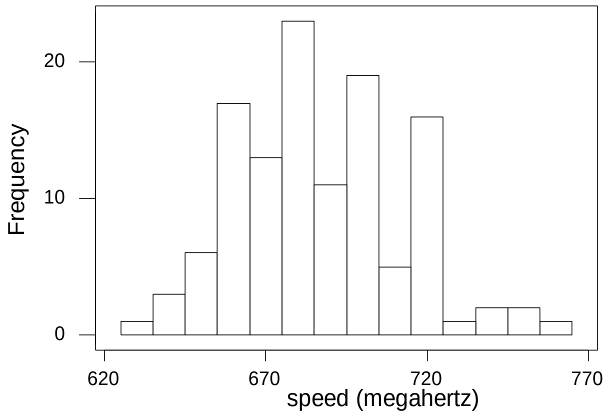


Yes, the histogram shows the same shape as the stem-and-leaf display.

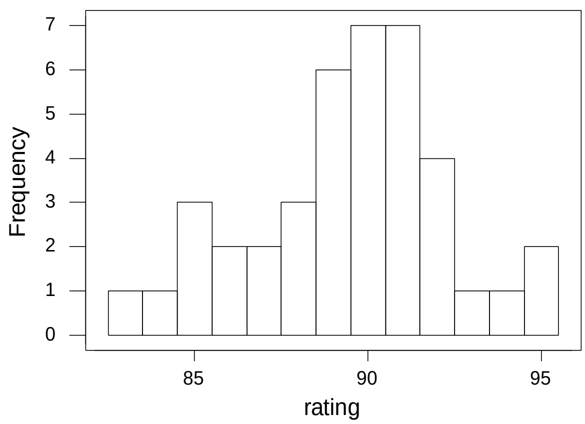
6-57. Yes, the histogram of the distance data shows the same shape as the stem-and-leaf display in exercise 6-33.



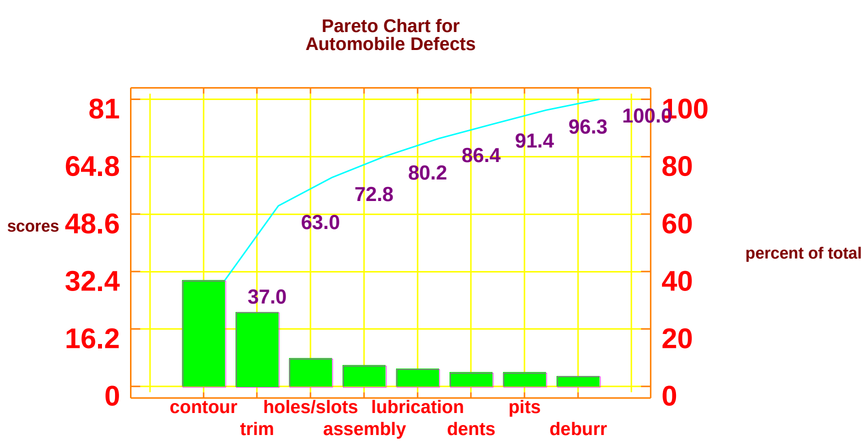
6-58. Histogram for the speed data. Yes, the histogram of the speed data shows the same shape as the stem-and-leaf display.



6-59. Yes, the histogram of the wine rating data shows the same shape as the stem-and-leaf display.



6-60.

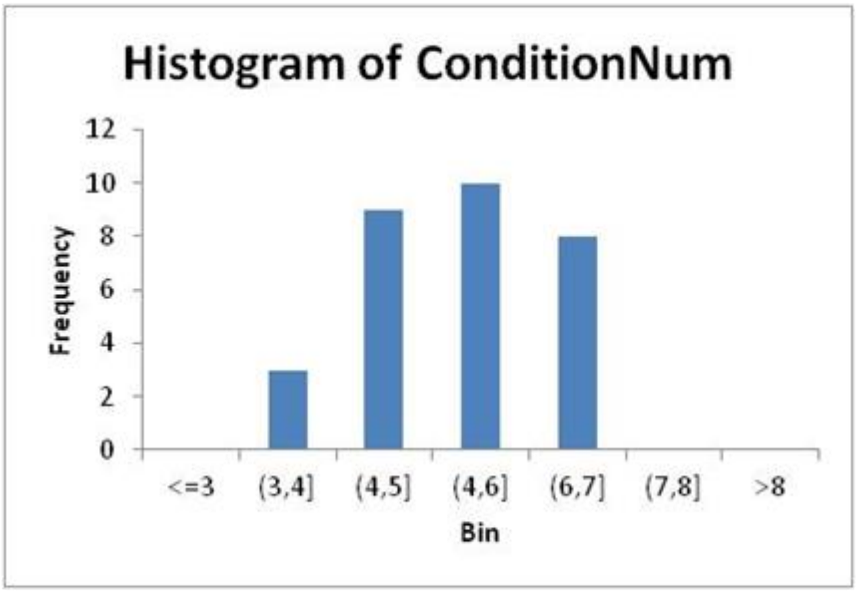


Roughly 63% of defects are described by parts out of contour and parts under trimmed.

6-61. Frequency Tabulation for Bridge Condition

Bin	*Frequency*
<=3	0
(3,4]	3
(4,5]	9
(4,6]	10
(6,7]	8
(7,8]	0
>8	0

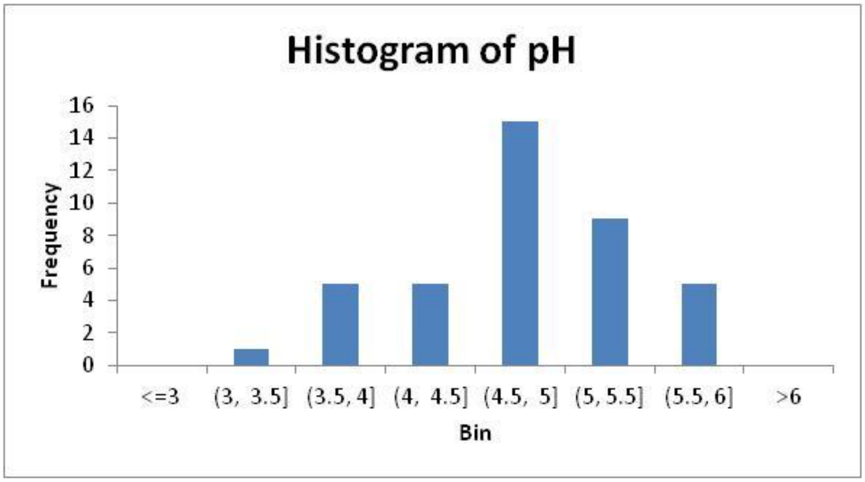
Mean	5.328333333
Median	5.22
Standard Deviation	0.975337548



6-62. Frequency Tabulation for Acid Rain

Bin	*Frequency*
<=3	0
(3, 3.5]	1
(3.5, 4]	5
(4, 4.5]	5
(4.5, 5]	15
(5, 5.5]	9
(5.5, 6]	5
>6	0

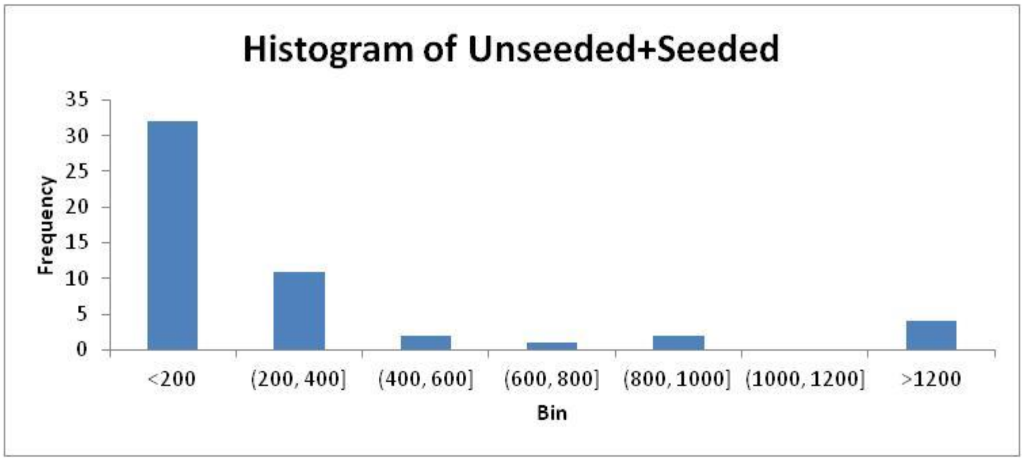
Mean	4.7205
Median	4.635
Standard Deviation	0.637144873



6-63. Frequency Tabulation for Cloud Seeding

Bin	*Frequency*
<200	32
(200, 400]	11
(400, 600]	2
(600, 800]	1
(800, 1000]	2
(1000, 1200]	0
>1200	4

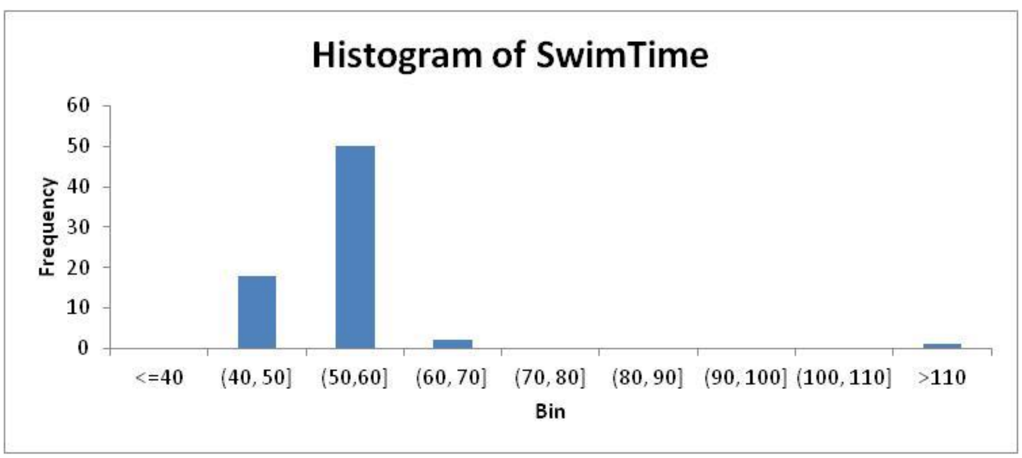
Mean	303.2865
Median	116.8
Standard Deviation	514.9998



6-64. Frequency Tabulation for Swim Times

Mean	52.64760563
Median	51.34
Standard Deviation	7.655977397

Bin	*Frequency*
<=40	0
(40, 50]	18
(50,60]	50
(60, 70]	2
(70, 80]	0
(80, 90]	0
(90, 100]	0
(100, 110]	0
>110	1



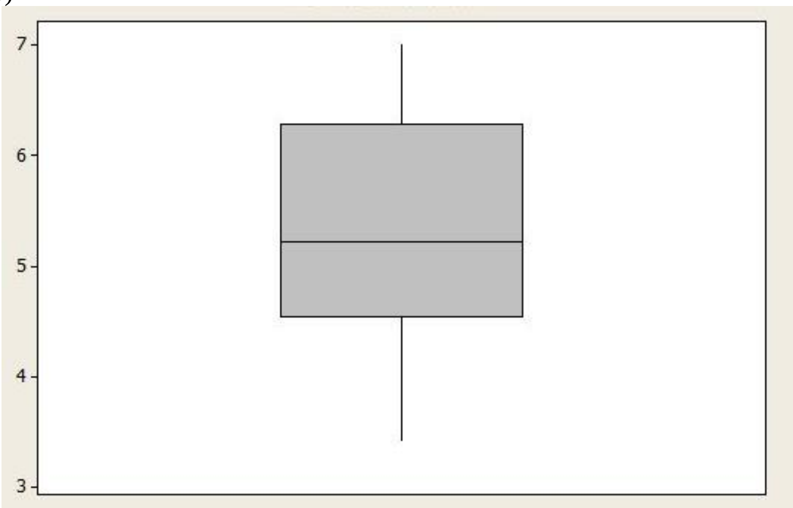
Section 6-4

6-65. a) Descriptive Statistics: Bridge Condition

```
Variable    N  N*    Mean  SE Mean  StDev  Minimum     Q1  Median     Q3
BrgCnd     30   0  5.328    0.178  0.975    3.420  4.538   5.220  6.277
Maximum
7.000
```

Median = 5.220
Lower Quartile = Q1= 4.538
Upper Quartile = Q3= 6.277

b)

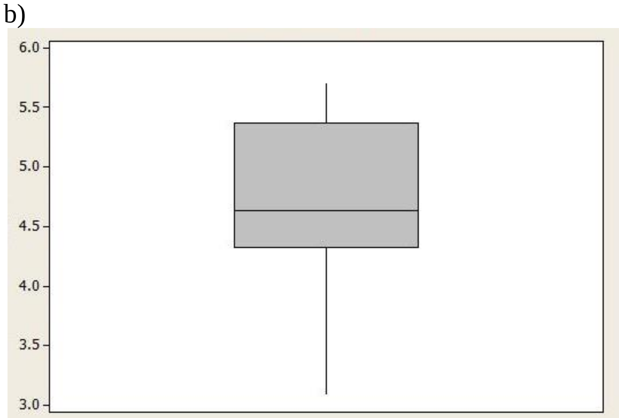


c) No obvious outliers.

6-66. a) Descriptive Statistics: Acid Rain

Variable	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
AcidRain	4.720	0.101	0.637	3.100	4.325	4.635	5.370	5.700

b)



c) No outliers are seen in the box plot whereas in the dot diagram a lower value is seen in the plot. The lower value in the dot diagram lies within 1.5 interquartile ranges of the first quartile in the box plot. Hence, it is not shown as an outlier in the box plot.

6-67. a) Descriptive Statistics for unseeded cloud data:

Variable	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Unseeded	164.6	54.6	278.4	1.0	23.7	44.2	183.3	1202.6

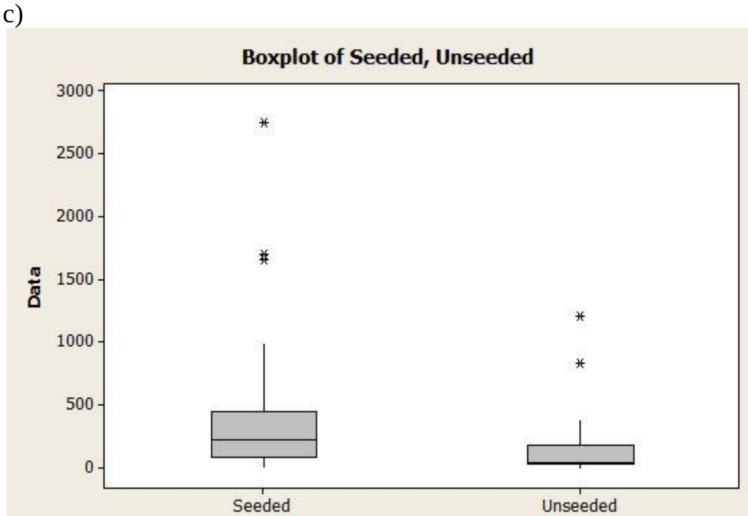
Median = 44.2
Lower Quartile = Q1 = 23.7
Upper Quartile = Q3 = 1202.6

b) Descriptive Statistics for seeded data:

Variable	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Seeded	442	128	651	4	79	222	445	2746

Median = 222
Lower Quartile = Q1 = 79
Upper Quartile = Q3 = 445

c)



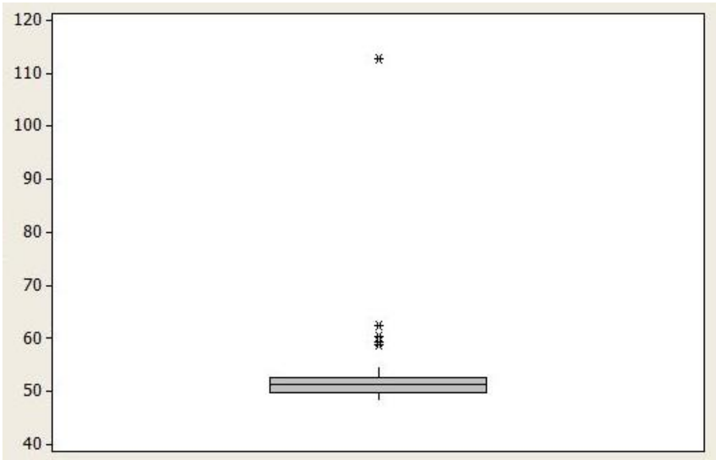
d) The two data sets are plotted here. A greater mean and greater dispersion in the seeded data are seen. Both plots show outliers.

6-68. a) Descriptive Statistics:

Variable	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Y	52.648	0.909	7.656	48.640	49.930	51.340	52.580	112.720

Median = 51.34
Lower Quartile = Q1 = 49.93
Upper Quartile = Q3 = 52.58

b)

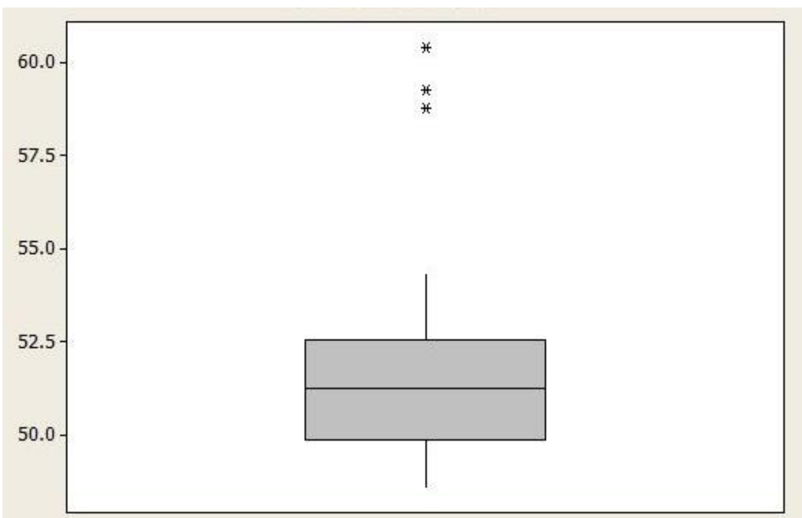


c) The extreme outliers present in this data are 62.45 and 112.72

Descriptive data without extreme outliers:

Variable	TotalCount	N	N*	Mean	SE Mean	StDev	Minimum	Q1
SwimTimes	71	69	2	51.635	0.264	2.194	48.640	49.885

Median	Q3	Maximum
51.280	52.550	60.390



d) Removing the extreme outliers reduced the mean and the median. Removing these outliers also substantially reduces the variability as seen by the smaller standard deviation and the smaller difference between the upper and lower quartiles.

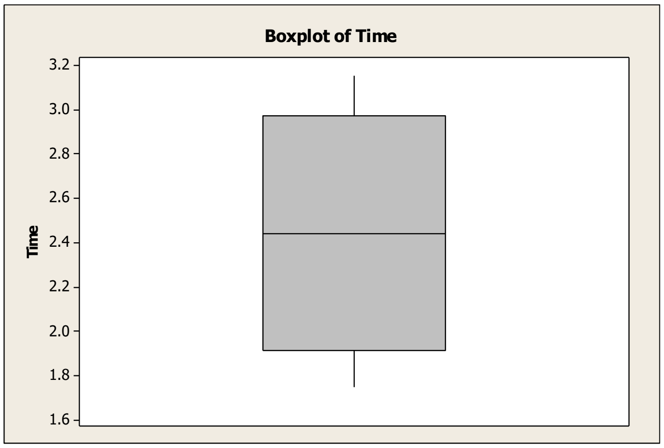
6-69. Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev	SE Mean
time	8	2.415	2.440	2.415	0.534	0.189

Variable	Minimum	Maximum	Q1	Q3
time	1.750	3.150	1.912	2.973

a)Sample Mean: 2.415, Sample Standard Deviation: 0.543

b) Box Plot: There are no outliers in the data.

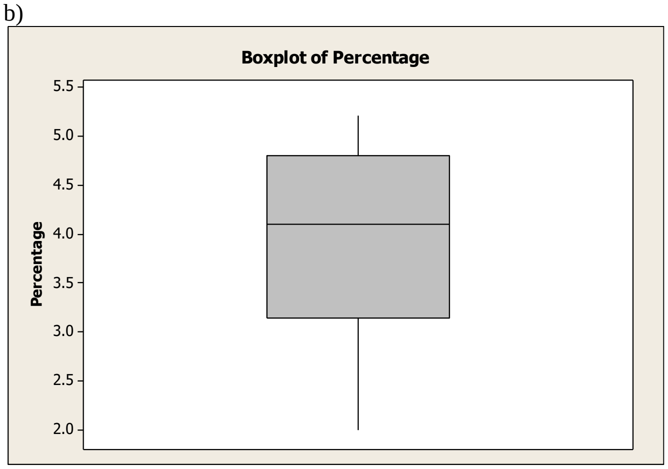


6-70. Descriptive Statistics

```
Variable        N     Mean   Median  Tr Mean    StDev  SE Mean
PMC            20    4.000    4.100    4.044    0.931    0.208
Variable      Min      Max       Q1       Q3
PMC         2.000    5.200    3.150    4.800
```

a) Sample Mean = 4, Sample Variance = 0.867, Sample Standard Deviation = 0.931

b)



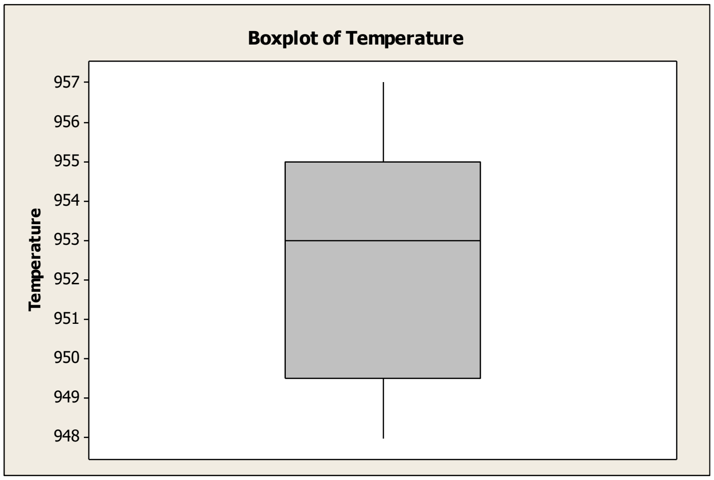
6-71. Descriptive Statistics

```
Variable        N     Mean   Median  Tr Mean    StDev  SE Mean
Temperat        9   952.44   953.00   952.44     3.09     1.03
Variable      Min      Max       Q1       Q3
Temperat   948.00   957.00   949.50   955.00
```

a) Sample Mean = 952.44, Sample Variance = 9.53, Sample Standard Deviation = 3.09

b) Median = 953; An increase in the largest temperature measurement does not affect the median.

c)

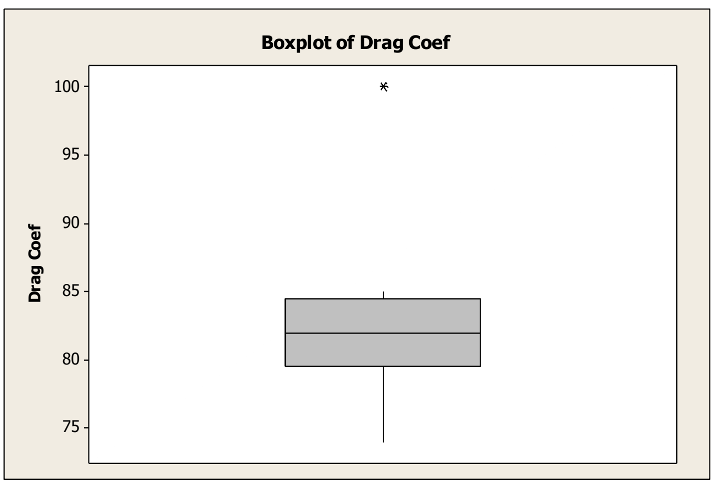


6-72. Descriptive statistics

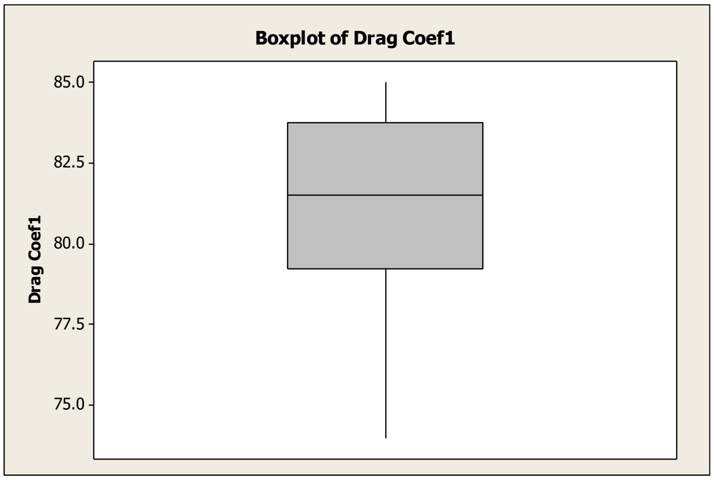
```
Variable              N   Mean   Median    TrMean    StDev   SE Mean
drag coefficients     9  83.11   82.00     83.11     7.11      2.37

Variable               Minimum   Maximum       Q1        Q3
drag coefficients        74.00    100.00    79.50     84.50
```

a) Median: $\tilde{x} = 82.00$, Upper quartile: Q_1 =79.50, Lower Quartile: Q_3=84.50

b)



c)
```
Variable              N   Mean   Median    TrMean    StDev   SE Mean
drag coefficients     8  81.00   81.50     81.00     3.46      1.22

Variable               Minimum   Maximum       Q1        Q3
drag coefficients        74.00     85.00    79.25     83.75
```



Removing the largest observation (100) decreases the mean and the median. Removing this "outlier" also greatly reduces the variability as seen by the smaller standard deviation and the smaller difference between the upper and lower quartiles.

6-73. Descriptive Statistics of O-ring joint temperature data

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Temp	36	65.86	67.50	66.66	12.16	2.03

Variable	Minimum	Maximum	Q1	Q3
Temp	31.00	84.00	58.50	75.00

a) Median = 67.50, Lower Quartile: $Q_1 = 58.50$, Upper Quartile: $Q_3 = 75.00$

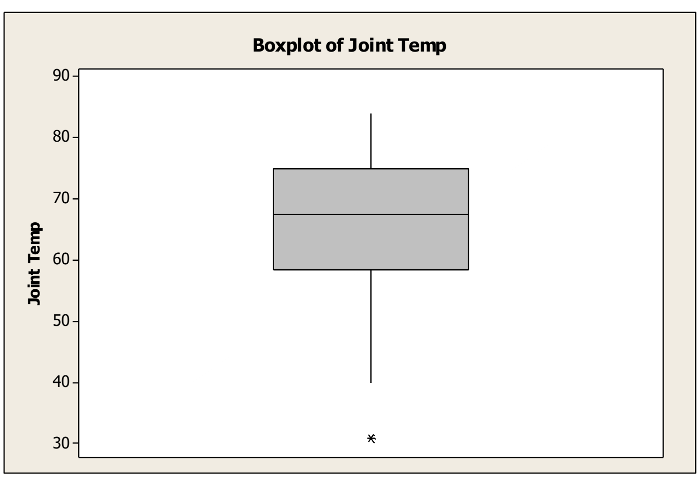
b) Data with lowest point removed

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Temp	35	66.86	68.00	67.35	10.74	1.82

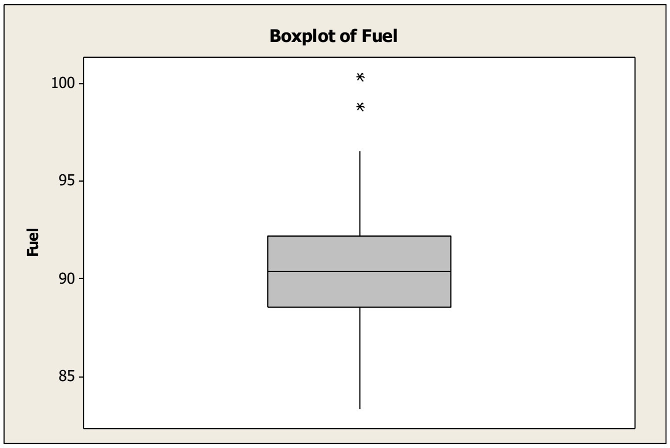
Variable	Minimum	Maximum	Q1	Q3
Temp	40.00	84.00	60.00	75.00

The mean and median have increased and the standard deviation and difference between the upper and lower quartile have decreased.

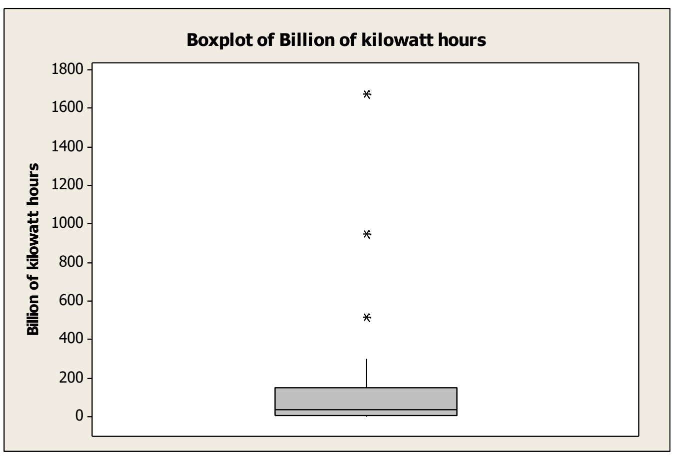
c)Box Plot: The box plot indicates that there is an outlier in the data.



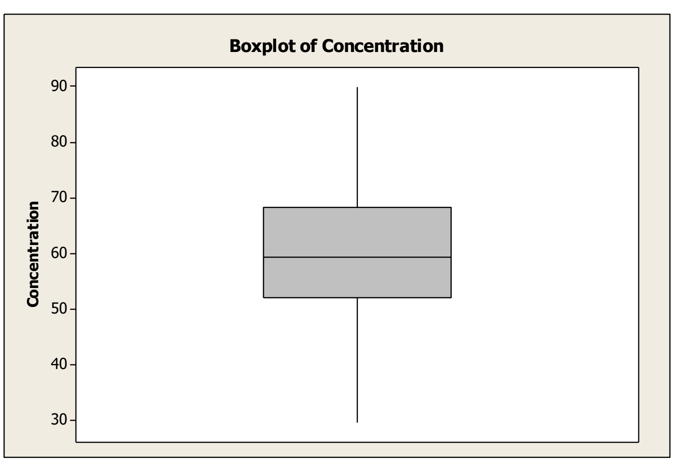
6-74. This plot conveys the same basic information as the stem and leaf plot but in a different format. The outliers that were separated from the main portion of the stem and leaf plot are shown here separated from the whiskers.



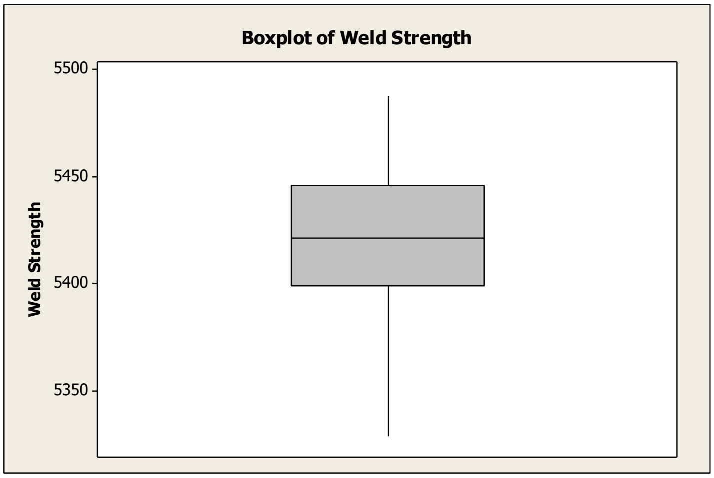
6-75. The box plot shows the same basic information as the stem and leaf plot but in a different format. The outliers that were separated from the main portion of the stem and leaf plot are shown here separated from the whiskers.



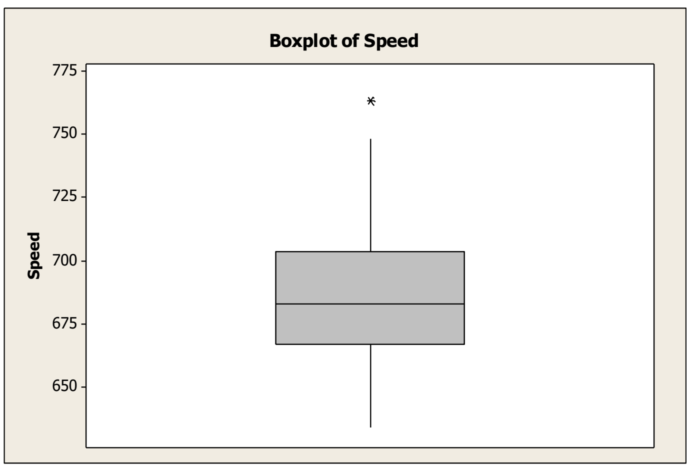
6-76. The box plot and the stem-leaf-diagram show that the data are very symmetrical about the mean. It also shows that there are no outliers in the data.



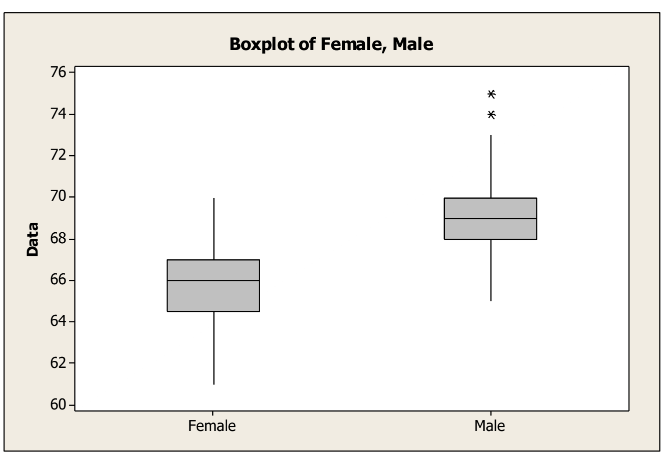
6-77. This plot, as the stem and leaf one, indicates that the data fall mostly in one region and that the measurements toward the ends of the range are more rare.



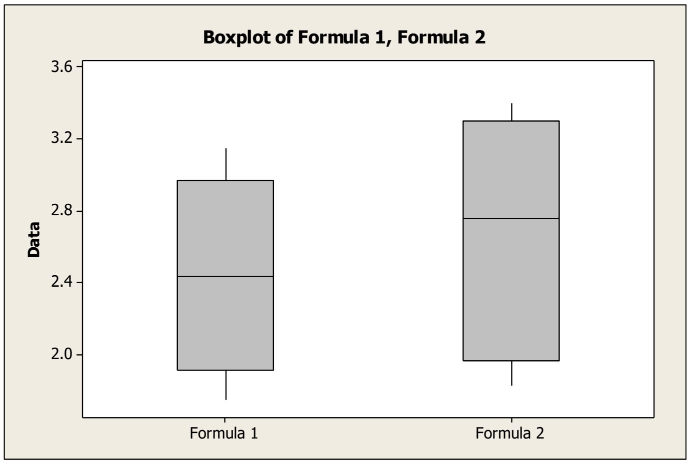
6-78. The box plot shows that the data are symmetrical about the mean. It also shows that there is an outlier in the data. These are the same interpretations seen in the stem-leaf-diagram.



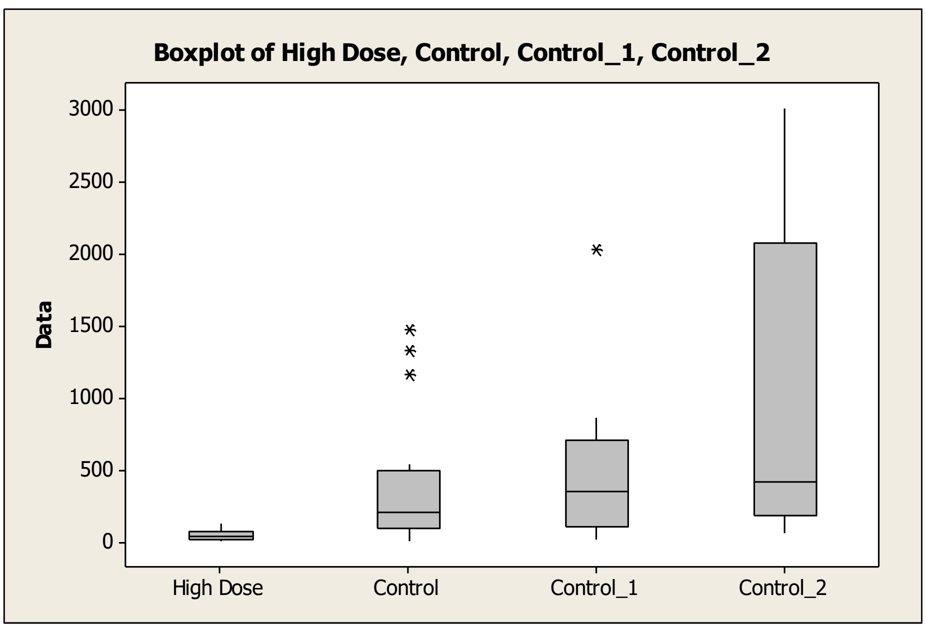
6-79. We can see that the two distributions seem to be centered at different values.



6-80. The box plot indicates that there is a difference between the two formulations. Formulation 2 has a higher mean cold start ignition time and a larger variability in the values of the start times. The first formulation has a lower mean cold start ignition time and is more consistent. Care should be taken though, because these box plots for formula 1 and formula 2 are based on only 8 and 10 data points, respectively. More data should be collected on each formulation.



6-81. All distributions are centered at about the same value, but have different variances.



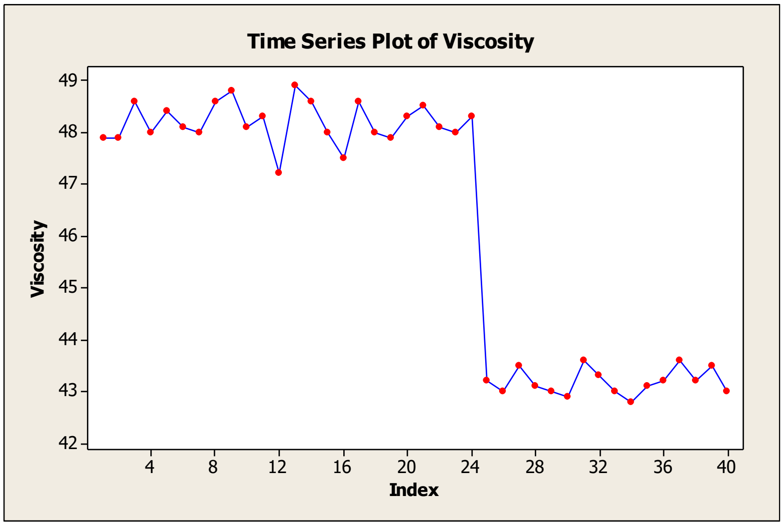
Section 6-5

6-82. Stem-leaf-plot of viscosity `N = 40`
`Leaf Unit = 0.10`

```
 2    42 89
12    43 0000112223
16    43 5566
16    44
16    44
16    45
16    45
16    46
16    46
17    47 2
(4)   47 5999
19    48 000001113334
 7    48 5666689
```

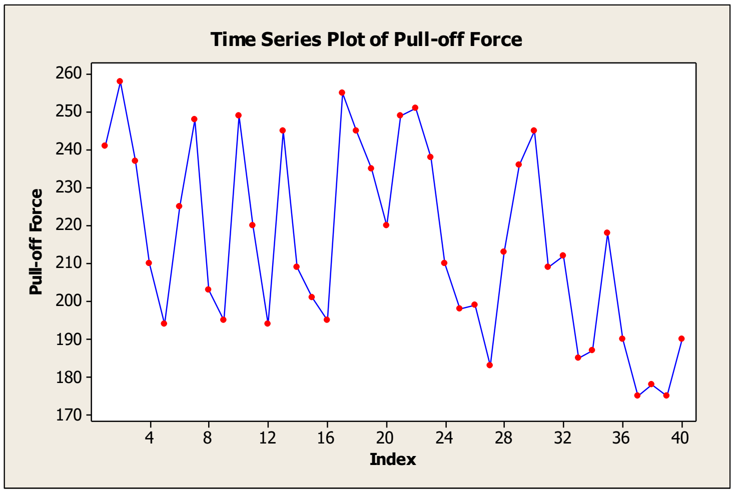
The stem-and-leaf plot shows that there are two different sets of data. One set of data is centered about 43 and the second set is centered about 48. The time series plot shows that the data starts out at the higher level and then drops down to the lower viscosity level at point 24. Each plot provides a different set of information.

If the specifications on the product viscosity are 48.0±2, then there is a problem with the process performance after data point 24. An investigation should take place to find out why the location of the process has dropped from around 48.0 to 43.0. The most recent product is not within specification limits.



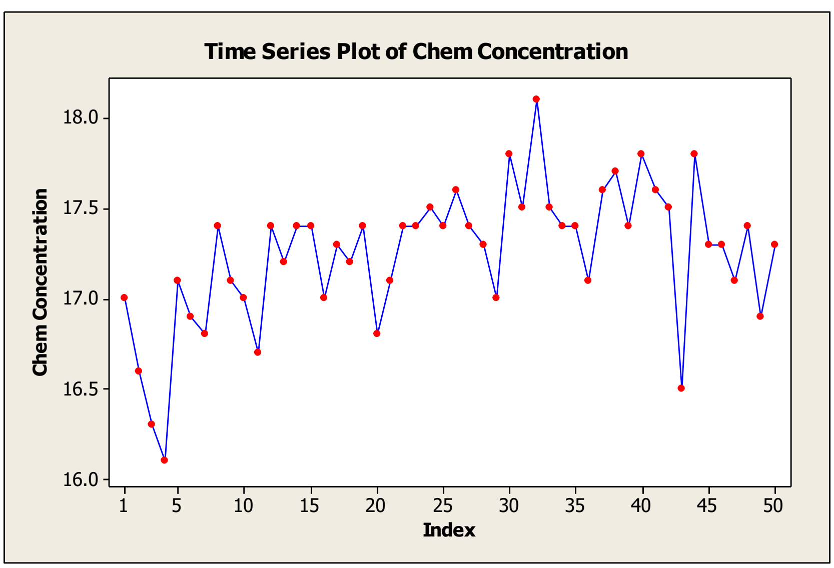
6-83. Stem-and-leaf display for Force: `unit = 1     1|2  represents 12`

```
 3    17|558
 6    18|357
14    19|00445589
18    20|1399
(5)   21|00238
17    22|005
14    23|5678
10    24|1555899
 3    25|158
```



In the time series plot there appears to be a downward trend beginning after time 30. The stem-and-leaf plot does not reveal this.

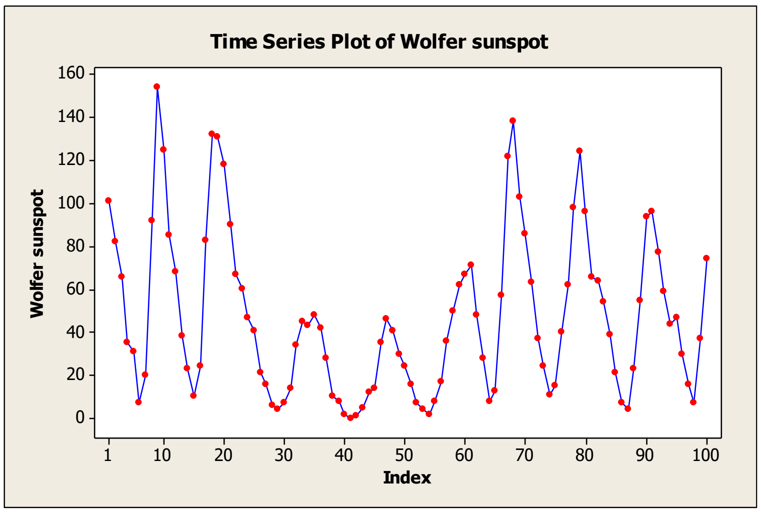
6-84.
```
Stem-and-leaf of Chem Concentration  N  = 50
Leaf Unit = 0.10


 1    16  1
 2    16  3
 3    16  5
 5    16  67
 9    16  8899
 18   17  000011111
 25   17  2233333
 25   17  44444444444445555
 8    17  6667
 4    17  888
 1    18  1
```



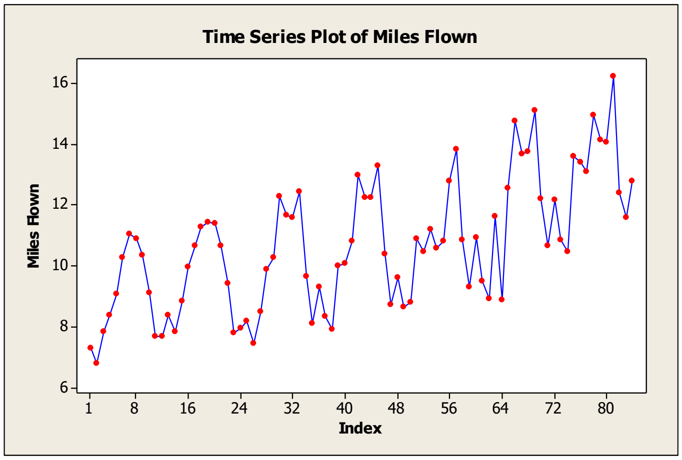
In the time series plot there appears to be trends with higher and lower concentration (probably autocorrelated data). The stem-and-leaf plot does not reveal this.

6-85.
```
Stem-and-leaf of Wolfer sunspot  N  = 100
Leaf Unit = 1.0

 17  0   01224445677777888
 29  1   001234456667
 39  2   0113344488
 50  3   00145567789
 50  4   011234567788
 38  5   04579
 33  6   0223466778
 23  7   147
 20  8   2356
 16  9   024668
 10  10  13
 8   11  8
 7   12  245
 4   13  128
 1   14
 1   15  4
```

The data appears to decrease between 1790 and 1835, the stem and leaf plot indicates skewed data.



6-86. Time Series Plot



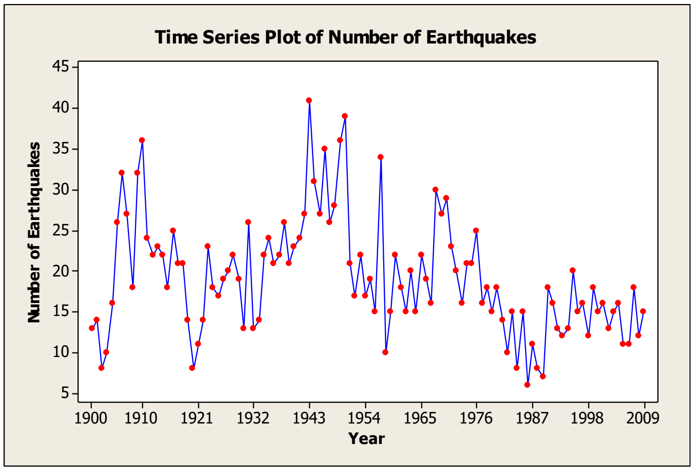
Each year the miles flown peaks during the summer hours. The number of miles flown increased over the years 1964 to 1970.

Stem-and-leaf of Miles Flown N = 84
```
Leaf Unit = 0.10


 1     6    7
 10    7    246678889
 22    8    013334677889
 33    9    01223466899
(18)   10   022334456667888889
 33    11   012345566
 24    12   11222345779
 13    13   1245678
 6     14   0179
 2     15   1
 1     16   2
```

When grouped together, the yearly cycles in the data are not seen. The data in the stem-leaf-diagram appear to be nearly normally distributed.

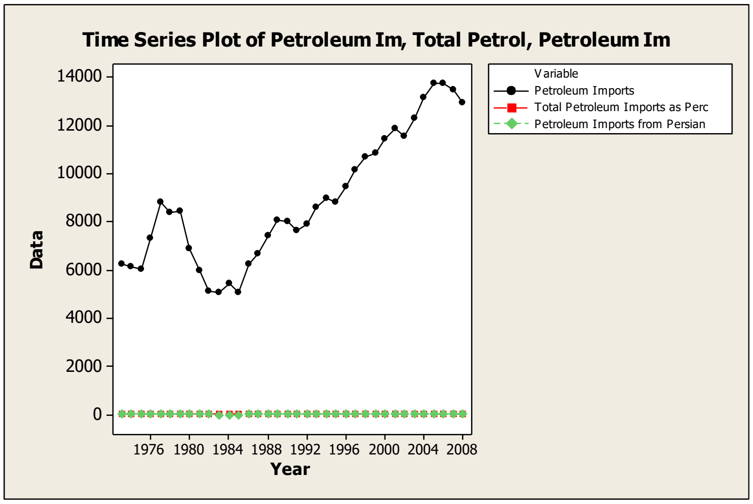
6-87.
```
Stem-and-leaf of Number of Earthquakes  N  = 110
Leaf Unit = 1.0

 2     0   67
 6     0   8888
 13    1   0001111
 22    1   222333333
 38    1   4444455555555555
 49    1   66666666777
(13)   1   8888888889999
 48    2   00001111111
 37    2   222222223333
 25    2   44455
 20    2   66667777
 12    2   89
 10    3   01
 8     3   22
 6     3   45
 4     3   66
 2     3   9
 1     4   1
```

Time Series Plot



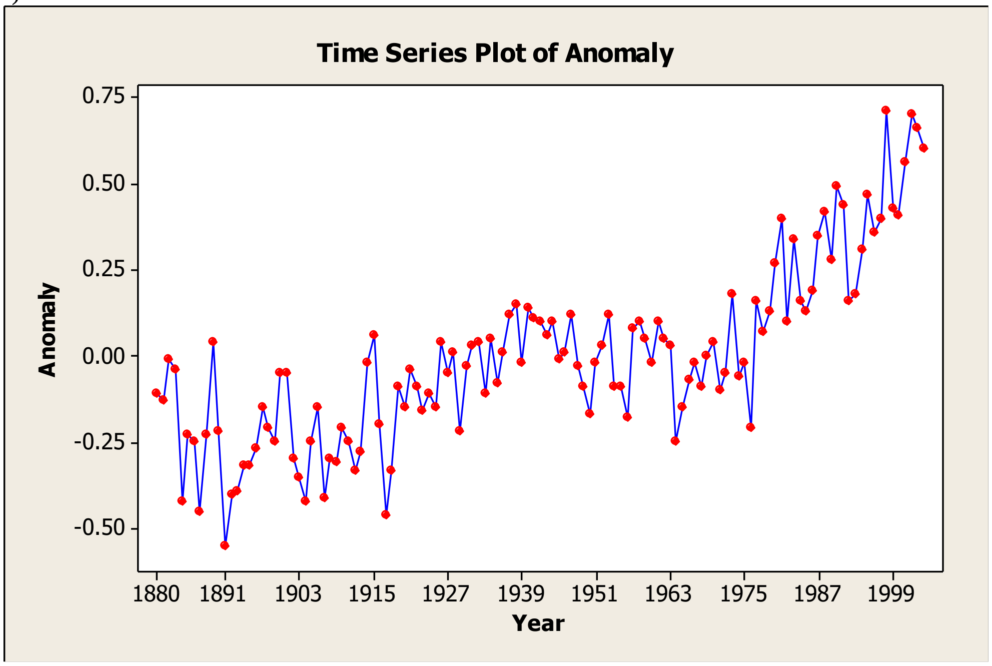
6-88.
```
Stem-and-leaf of Petroleum Imports  N  = 36
Leaf Unit = 100

 5    5   00149
 11   6   012269
 15   7   3468
(8)   8   00346889
 13   9   4
 12   10  178
 9    11  458
 6    12  29
 4    13  1477


Stem-and-leaf of Total Petroleum Imports as Perc  N  = 36
Leaf Unit = 1.0

 4    3  2334
 9    3  66778
 14   4  00124
(7)   4  5566779
 15   5  0014
 11   5  5688
 7    6  013
 4    6  5566


Stem-and-leaf of Petroleum Imports from Persian  N  = 36
Leaf Unit = 1.0

 1    0  6
 3    0  89
 3    1
 5    1  33
 6    1  4
 14   1  66667777
(6)   1  889999
 16   2  000011
 10   2  2233
 6    2  4445
 2    2  67
```

Time Series plot:



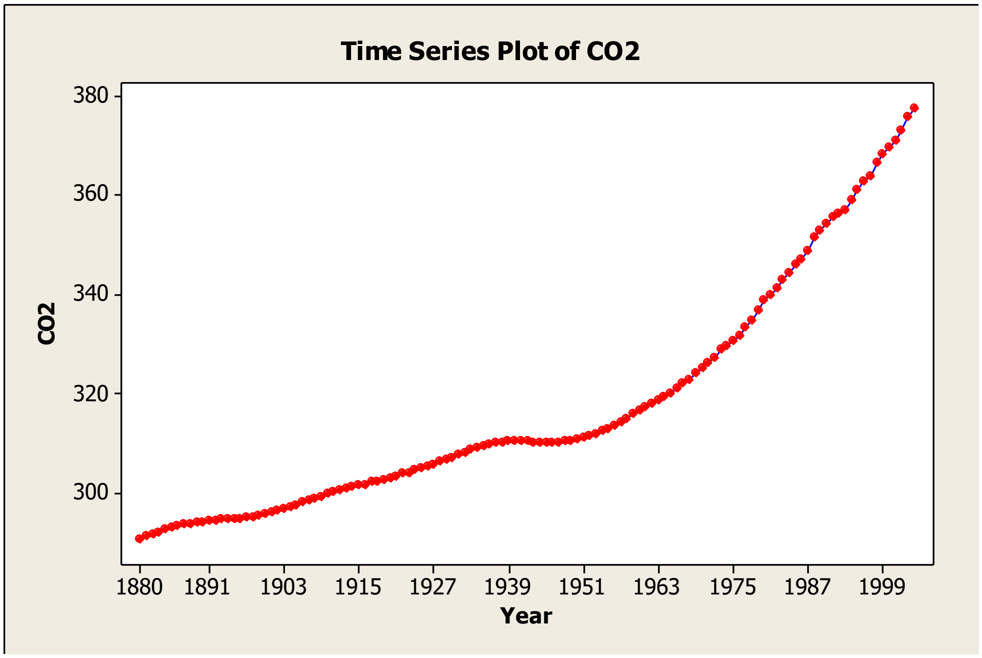
6-89.

a)



There is an increasing trend in the most recent data.

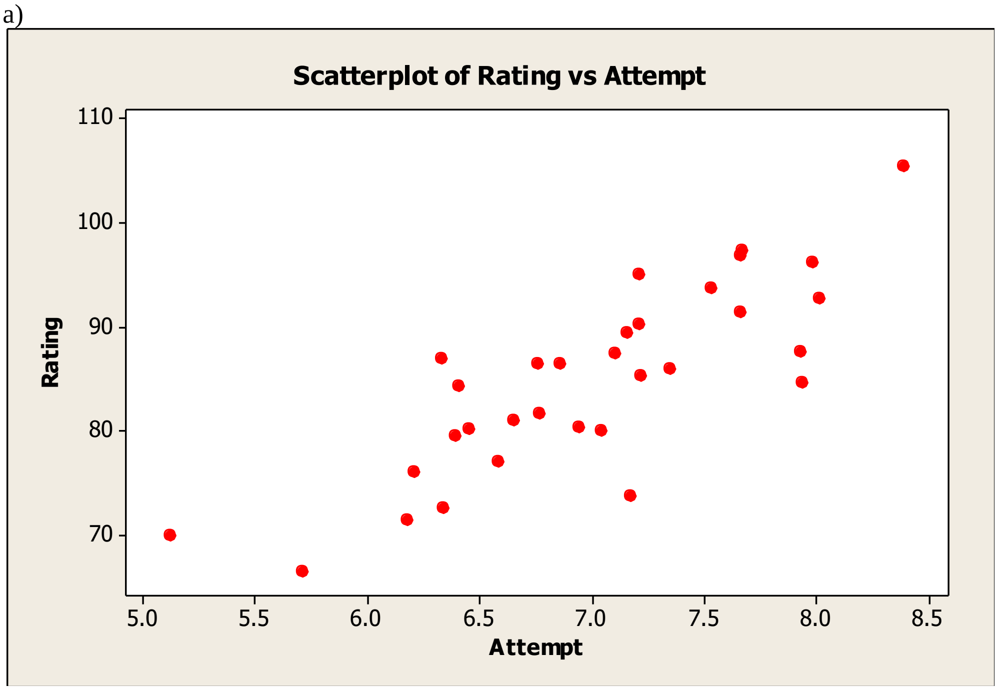
b)



c) The plots increase approximately together. However, this relationship alone does not prove a cause and effect.

Section 6-6

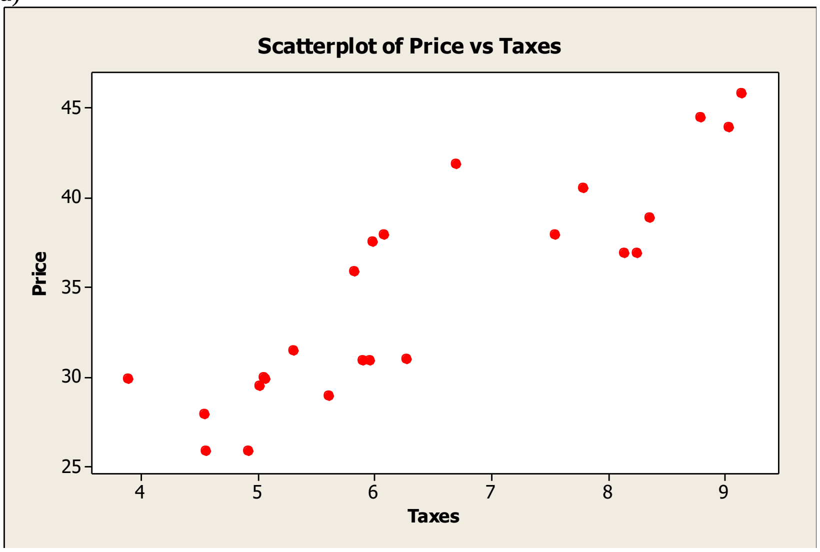
6-90. a)



As the *yards per attempt* increase, the *rating* tends to increase.

b) The correlation coefficient from computer software is 0.820

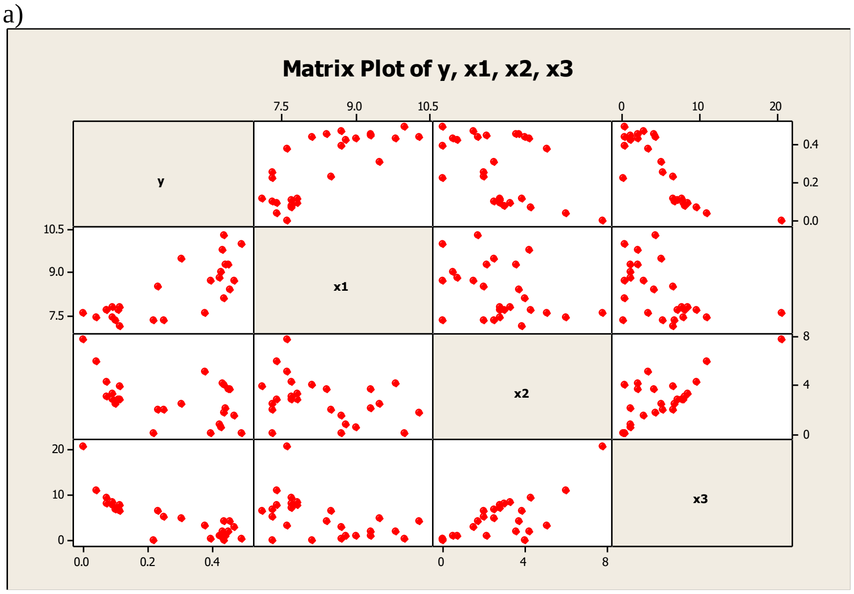
6-91. a)



As the *taxes* increase, the *price* tends to increase.

b) From computer software the correlation coefficient is 0.876

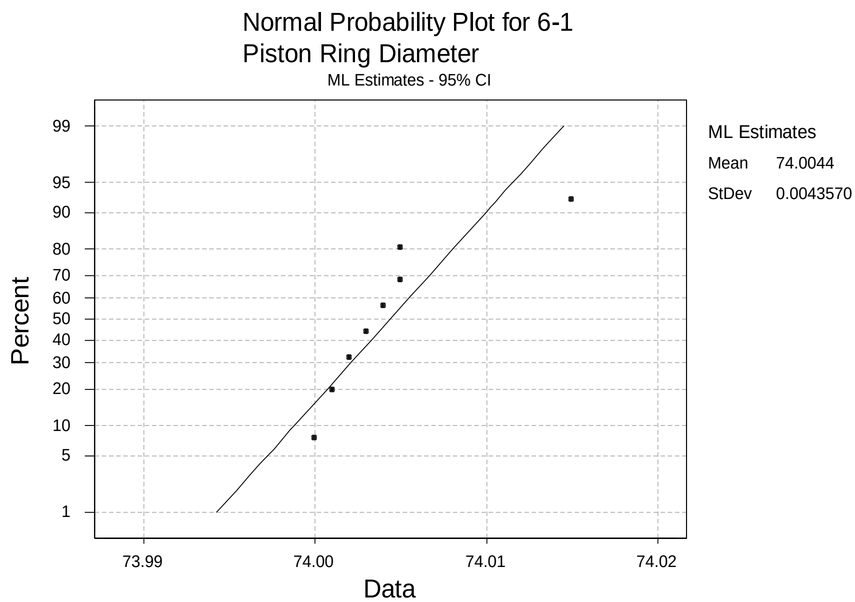
6-92. a)



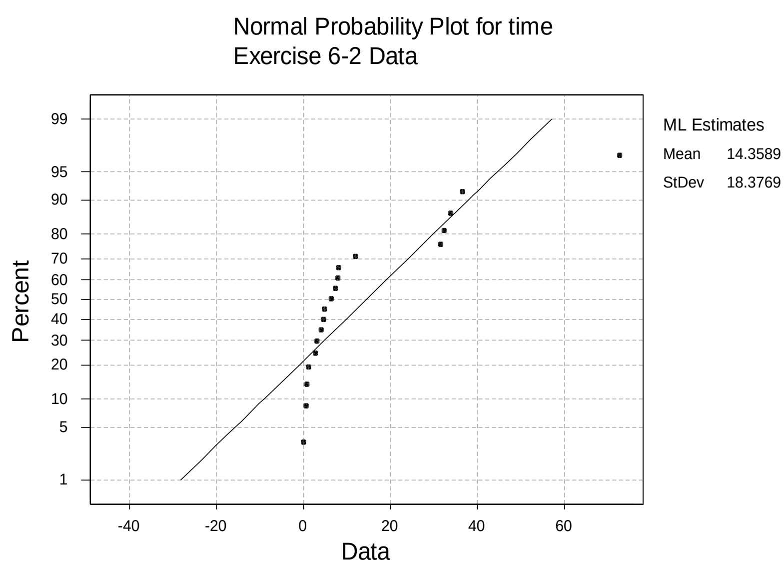
b) Values for y tend to increase as values for x_1 increase. However, values for y tend to decrease as values for x_2 or x_3 decrease.

6-93.

The pattern of the data indicates that the sample may not come from a normally distributed population or that the largest observation is an outlier. Note the slight bending downward of the sample data at both ends of the graph.

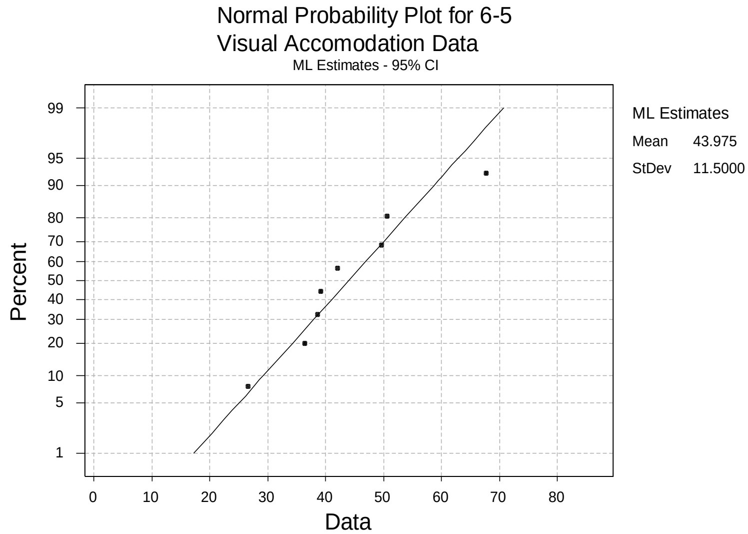


6-94. It appears that the data do not come from a normal distribution. Very few of the data points fall near the line.

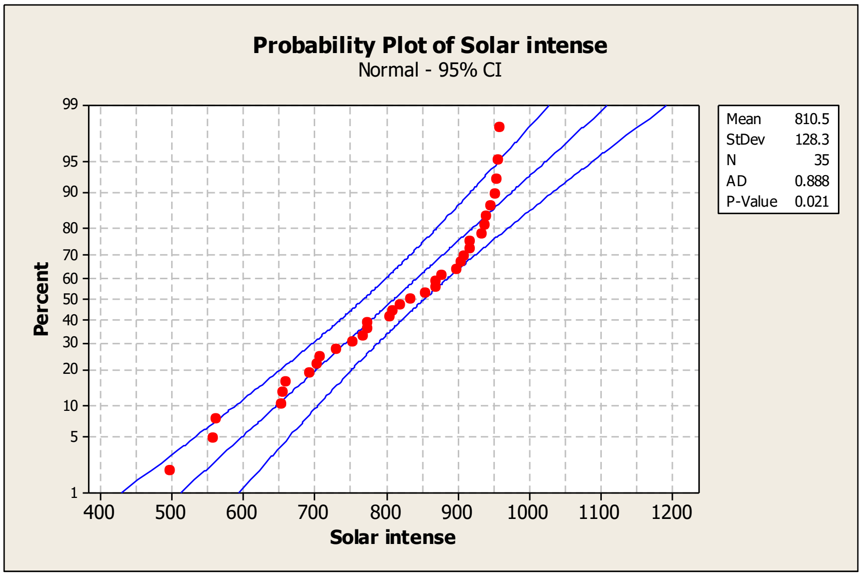


6-95.

A normal distribution is reasonable for these data.

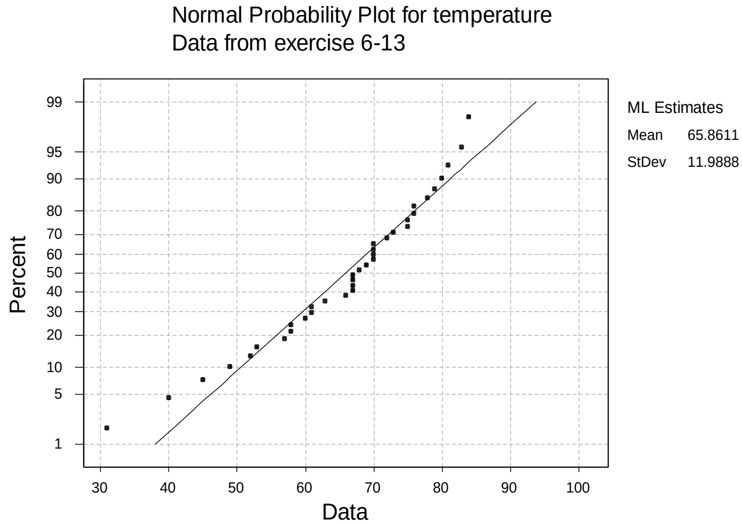


6-96. The normal probability plot shown below does not seem reasonable for normality.

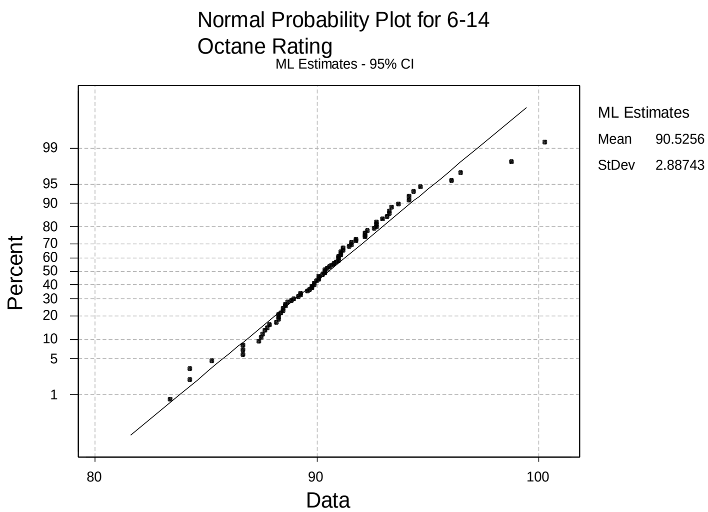


6-97.

The data appear to be approximately normally distributed. However, there are some departures from the line at the ends of the distribution.

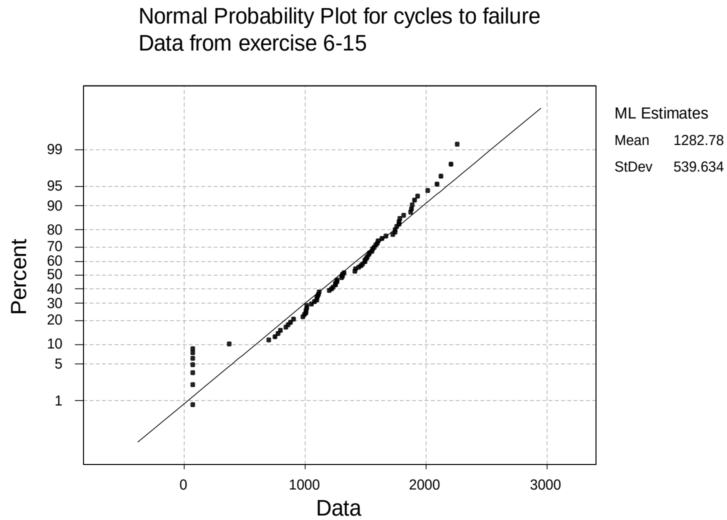


6-98.



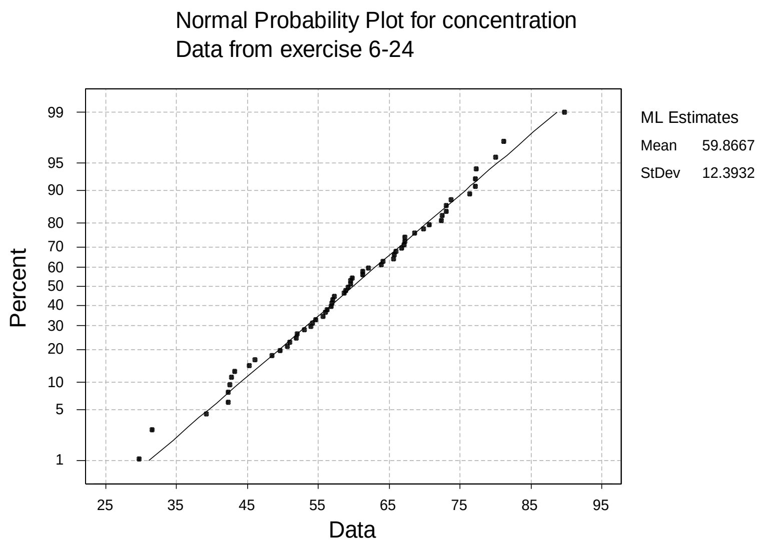
The data appear to be approximately normally distributed. However, there are some departures from the line at the ends of the distribution.

6-99.



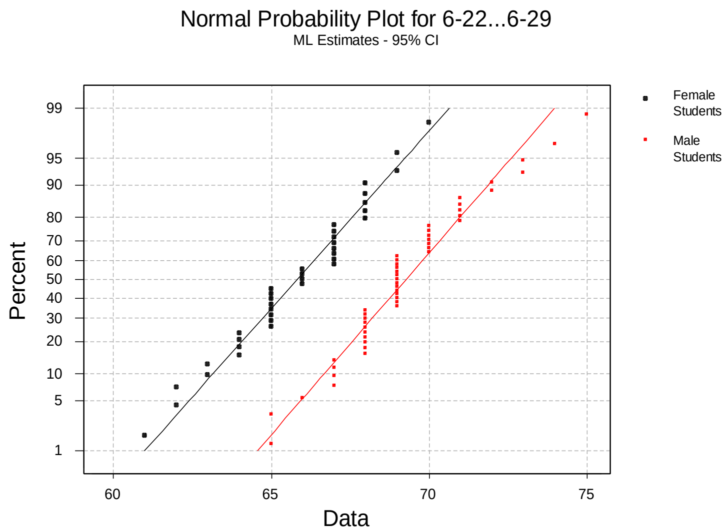
The data appear to be approximately normally distributed. However, there are some departures from the line at the ends of the distribution.

6-100.



The data appear to be normally distributed. Nearly all of the data points fall very close to the line.

6-101.

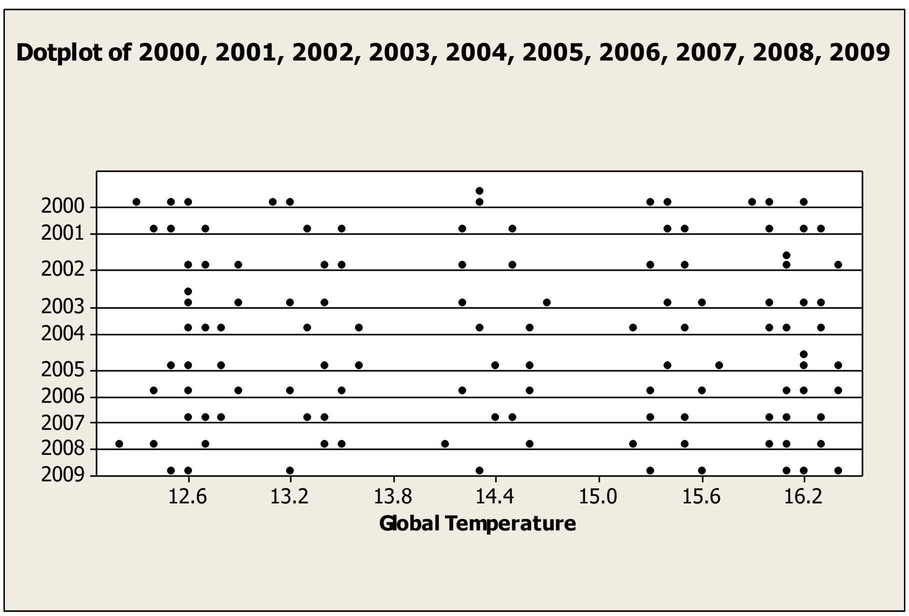


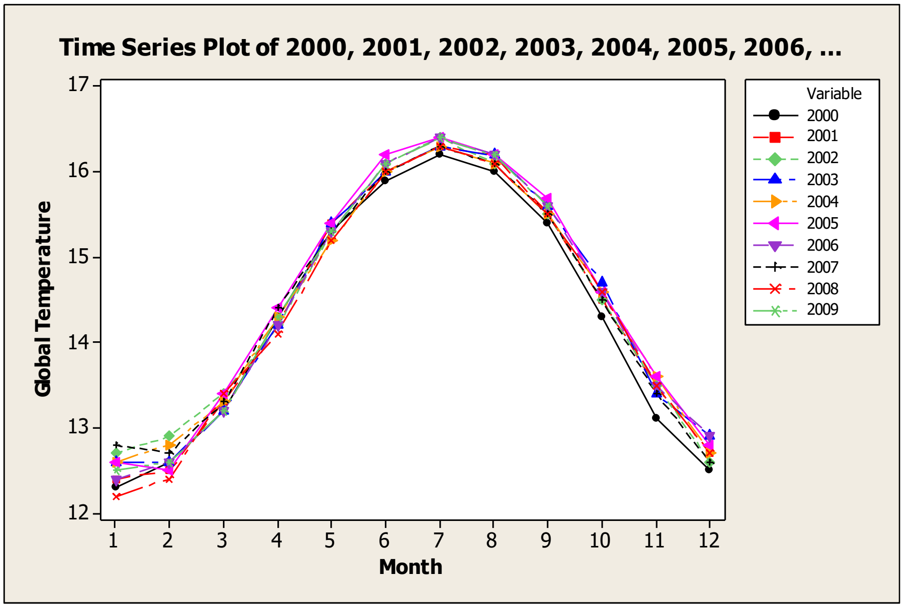
Both populations seem to be normally distributed. Moreover, the lines seem to be roughly parallel indicating that the populations may have the same variance and differ only in the value of their mean.

6-102. Yes, it is possible to obtain an estimate of the mean from the 50th percentile value of the normal probability plot. The 50th percentile point is the median, and for a normal distribution the median equals the mean. An estimate of the standard deviation can be obtained from the 84th percentile minus the 50th percentile . From the z-table, the 84th percentile of a normal distribution is one standard deviation above the mean.

<u>Supplemental Exercises</u>

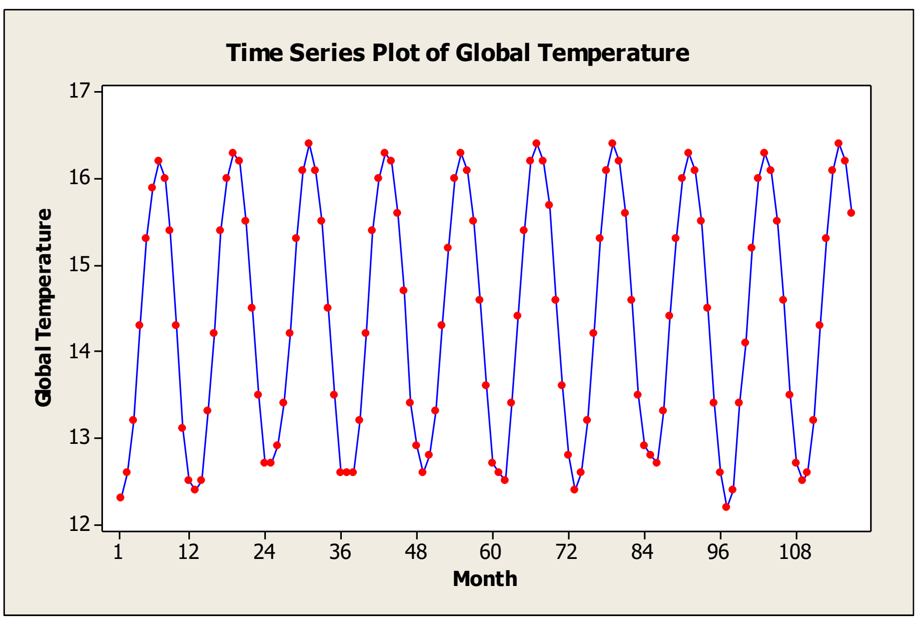
6-103. Based on the digidot plot and time series plots of these data, in each year the temperature has a similar distribution. In each year, the temperature increases until the mid year and then it starts to decrease.






```
Stem-and-leaf of Global Temperature  N  = 117
Leaf Unit = 0.10


 5    12  23444
 29   12  555566666666677777888999
 42   13  1222233344444
 48   13  555566
(11)  14  12222333344
 58   14  55566667
 50   15  22333334444
 39   15  5555566679
 29   16  00000011111112222222333334444
```

Time-series plot by month over 10 years



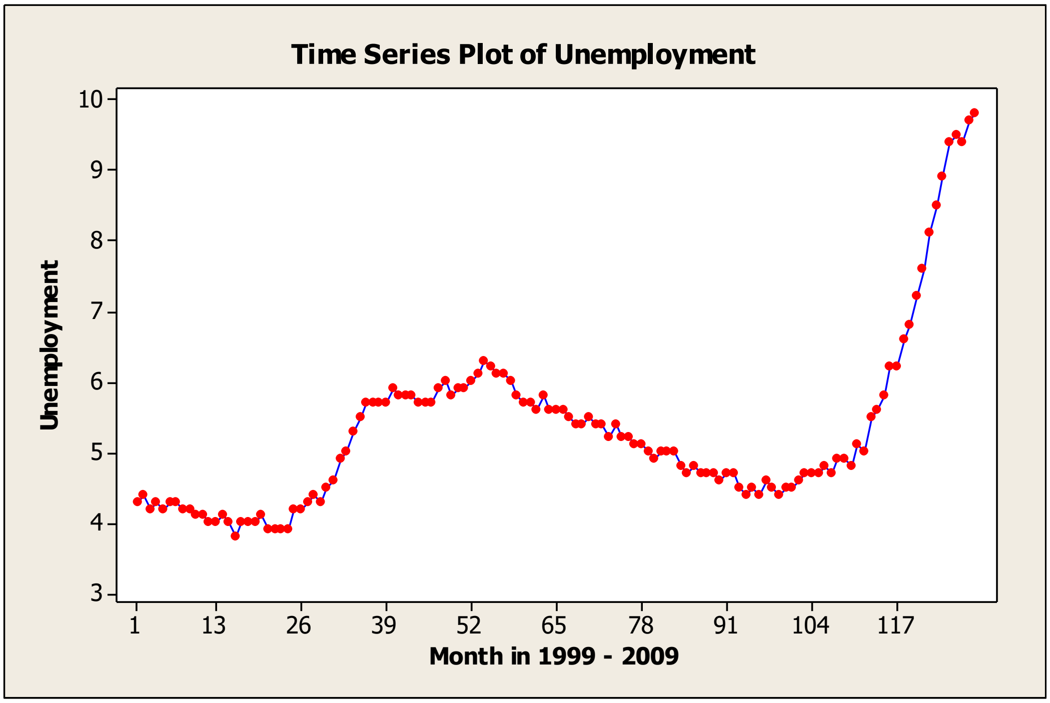
6-104. a) Sample Mean = 65.083
The sample mean value is close enough to the target value to accept the solution as conforming. There is a slight difference due to inherent variability.

b) $s^2 = 1.86869$ $s = 1.367$

c) A major source of variability might be variability in the reagent material. Furthermore, if the same setup is used for all measurements it is not expected to affect the variability. However, if each measurement uses a different setup, then setup differences could also be a major source of variability.

A low variance is desirable because it indicates consistency from measurement to measurement. This implies the measurement error has low variability.

6-105.
The unemployment rate is steady from 200-2002, then it increases until 2004, decreases steadily from 2004 to 2008, and then increases again dramatically in 2009, where it peaks.



6-106. a) $\sum_{i=1}^{6} x_i^2 = 10{,}433$ $\left(\sum_{i=1}^{6} x_i\right)^2 = 62{,}001$ $n = 6$

$$s^2 = \frac{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{n}}{n-1} = \frac{10{,}433 - \frac{62{,}001}{6}}{6-1} = 19.9\Omega^2$$

$$s = \sqrt{19.9\Omega^2} = 4.46\Omega$$

b) $\sum_{i=1}^{6} x_i^2 = 353$ $\left(\sum_{i=1}^{6} x_i\right)^2 = 1521$ $n = 6$

$$s^2 = \frac{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{n}}{n-1} = \frac{353 - \frac{1{,}521}{6}}{6-1} = 19.9\Omega^2$$

$$s = \sqrt{19.9\Omega^2} = 4.46\Omega$$

Shifting the data from the sample by a constant amount has no effect on the sample variance or standard deviation.

c) $\sum_{i=1}^{6} x_i^2 = 1043300$ $\left(\sum_{i=1}^{6} x_i\right)^2 = 6200100$ $n = 6$

$$s^2 = \frac{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{n}}{n-1} = \frac{1043300 - \frac{6200100}{6}}{6-1} = 1990\Omega^2$$

$$s = \sqrt{1990\Omega^2} = 44.61\Omega$$

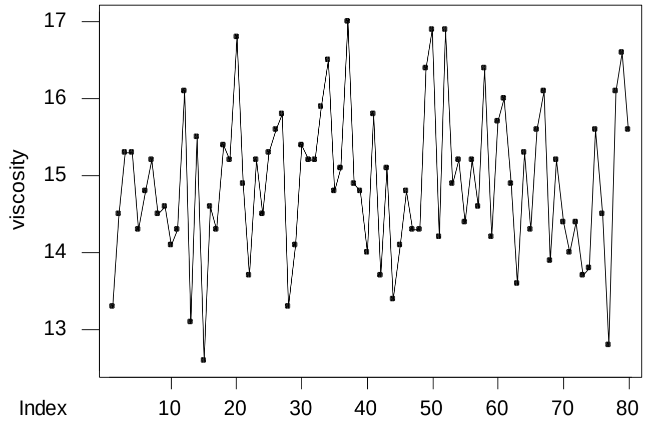
Yes, the rescaling is by a factor of 10. Therefore, s^2 and s would be rescaled by multiplying s^2 by 10^2 (resulting in $1990\Omega^2$) and s by 10 (44.6Ω).

6-107. a) Sample 1 Range = 4, Sample 2 Range = 4
Yes, the two appear to exhibit the same variability

b) Sample 1 s = 1.604, Sample 2 s = 1.852
No, sample 2 has a larger standard deviation.

c) The sample range is a relatively crude measure of the sample variability as compared to the sample standard deviation because the standard deviation uses the information from every data point in the sample whereas the range uses the information contained in only two data points - the minimum and maximum.

6-108. a) It appears that the data may shift up and then down over the 80 points.



b) It appears that the mean of the second set of 40 data points may be slightly higher than the first set of 40.

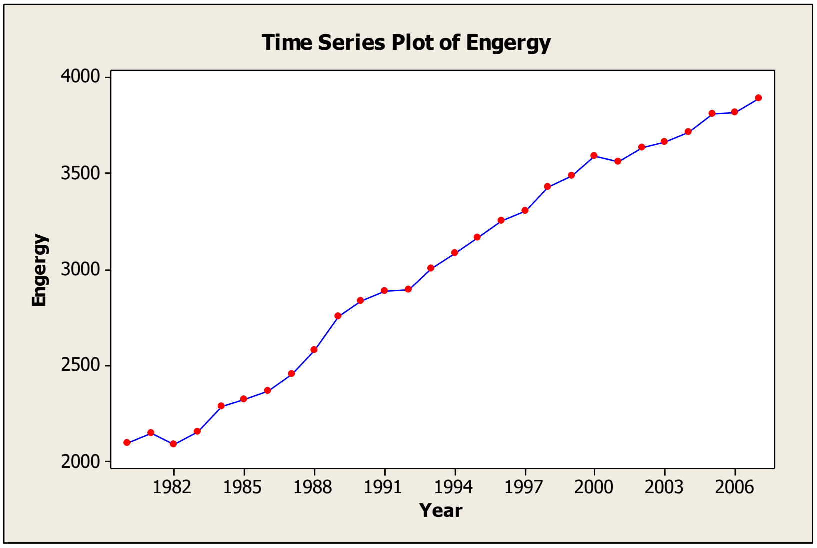
c) Descriptive Statistics: viscosity 1, viscosity 2

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Viscosity1	40	14.875	14.900	14.875	0.948	0.150
Viscosity2	40	14.923	14.850	14.914	1.023	0.162

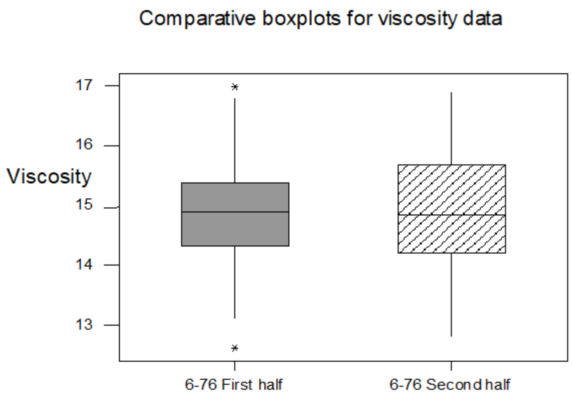
There is a slight difference in the mean levels and the standard deviations.

6-109. From the stem-and-leaf diagram, the distribution looks like the uniform distribution. From the time series plot, there is an increasing trend in energy consumption.

```
Stem-and-leaf of Energy  N  = 28
Leaf Unit = 100

 4   2  0011
 7   2  233
 9   2  45
 10  2  7
 13  2  888
(3)  3  001
 12  3  23
 10  3  4455
 6   3  667
 3   3  888
```

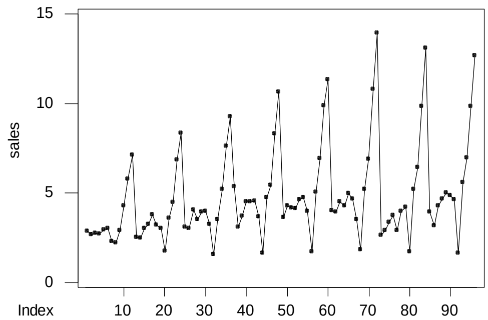


6-110.



Both sets of data appear to have the same mean although the first half of the data seems to be concentrated a little more tightly. Two data points appear as outliers in the first half of the data.

6-111.



There appears to be a cyclic variation in the data with the high value of the cycle generally increasing. The high values are during the winter holiday months.

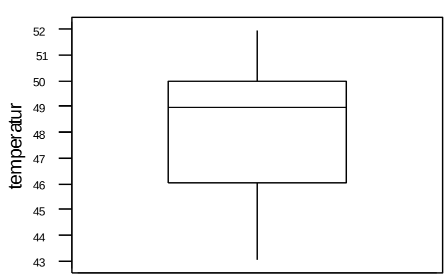
b) We might draw another cycle, with the peak similar to the last year's data (1969) at about 12.7 thousand bottles.

6-112.

Descriptive Statistics

```
Variable         N      Mean    Median  Tr Mean     StDev  SE Mean
temperat        24    48.125    49.000   48.182     2.692    0.549
Variable       Min       Max        Q1       Q3
temperat    43.000    52.000    46.000   50.000
```

a) Sample Mean: 48.12, Sample Median: 49
b) Sample Variance: 7.246, Sample Standard Deviation: 2.692
c)



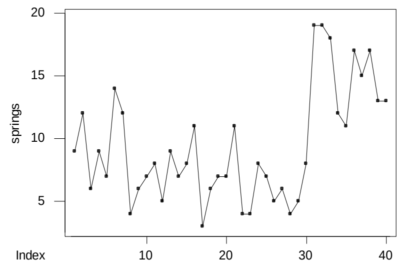
The data appear to be slightly skewed.

6-113. a) Stem-and-leaf display for Problem 2-35: unit = 1   1|2 represents 12

```
  1     0T|3
  8     0F|4444555
 18     0S|6666777777
 (7)    0o|8888999
 15     1*|111
 12     1T|22233
  7     1F|45
  5     1S|77
  3     1o|899
```

b) Sample Average = 9.325, Sample Standard Deviation = 4.4858

c)

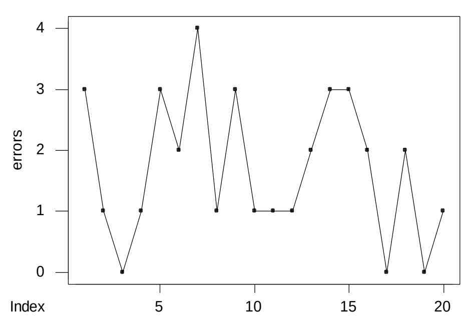


The time series plot indicates there was an increase in the average number of nonconforming springs during the 40 days. In particular, the increase occurred during the last 10 days.

6-114. a)
```
Stem-and-leaf of errors    N  = 20
  Leaf Unit = 0.10


   3    0 000
  10    1 0000000
  10    2 0000
   6    3 00000
   1    4 0
```

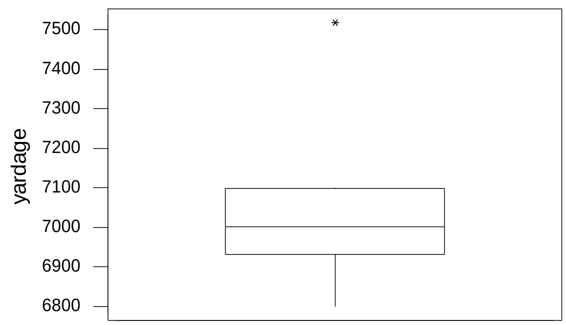
b) Sample Average = 1.700
Sample Standard Deviation = 1.174

c)



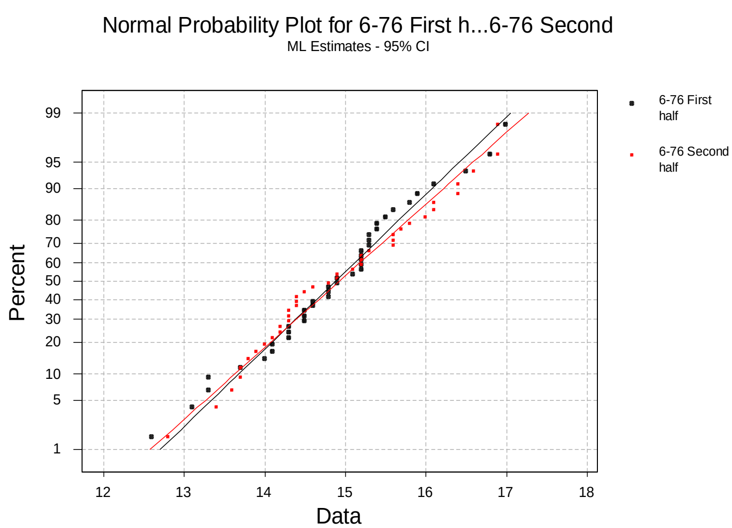
The time series plot indicates a slight decrease in the number of errors for strings 16 - 20.

6-115. The golf course yardage data appear to be skewed. Also, there is an outlying data point above 7500 yards.

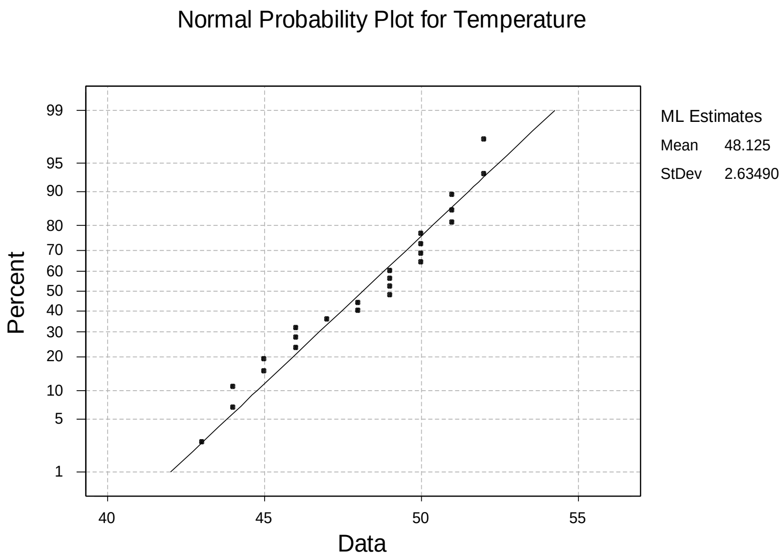


6-116.

Both sets of data appear to be normally distributed and with roughly the same mean value. The difference in slopes for the two lines indicates that a change in variance might have occurred. This could have been the result of a change in processing conditions, the quality of the raw material or some other factor.

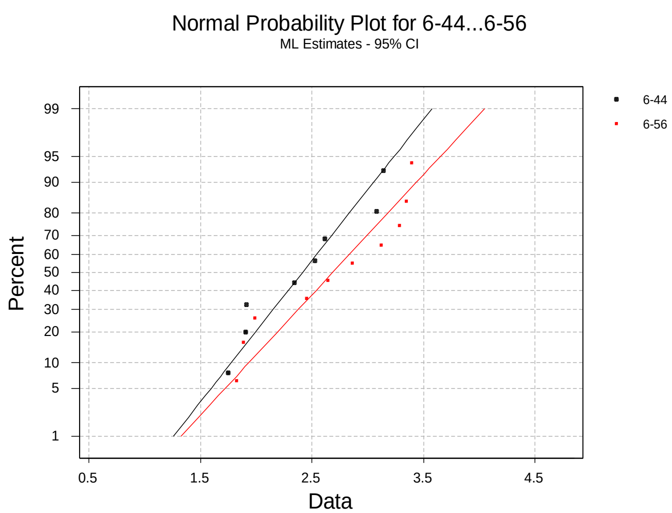


6-117.



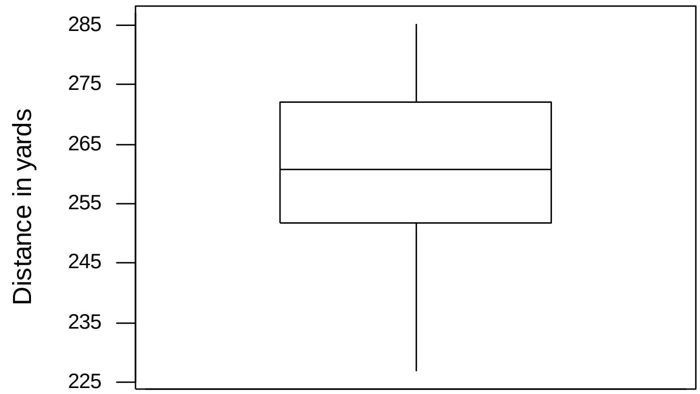
A normal distribution is reasonable for these data. There are some repeated values in the data that cause some points to fall off the line.

6-118.



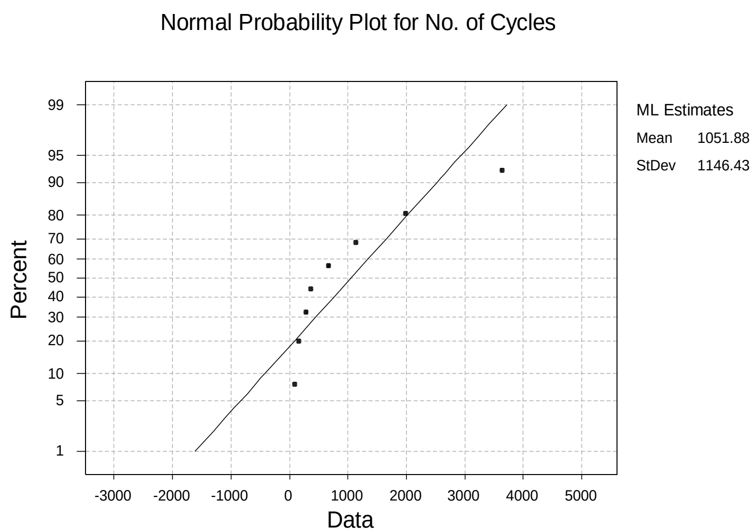
Although we do not have sufficient data points to really see a pattern, there seem to be no significant deviations from normality for either sample. The large difference in slopes indicates that the variances of the populations are very different.

6-119.



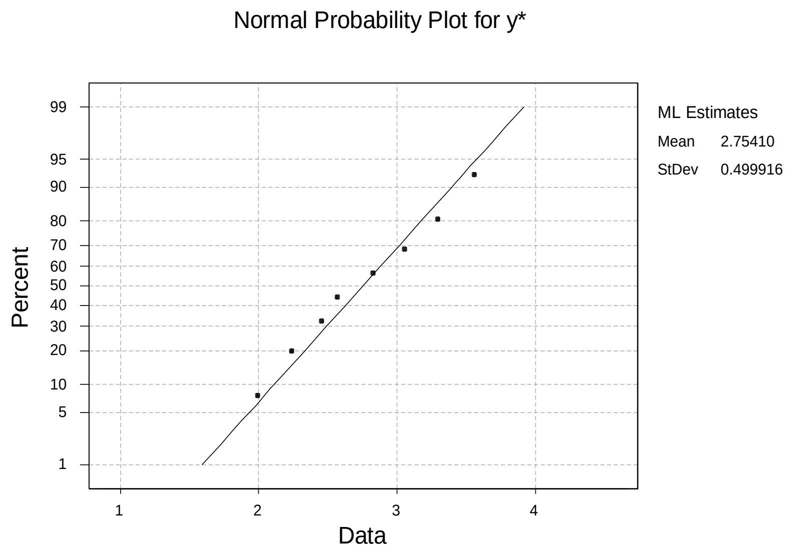
The plot indicates that most balls will fall somewhere in the 250-275 range. This same type of information could have been obtained from the stem and leaf graph.

6-120. a)



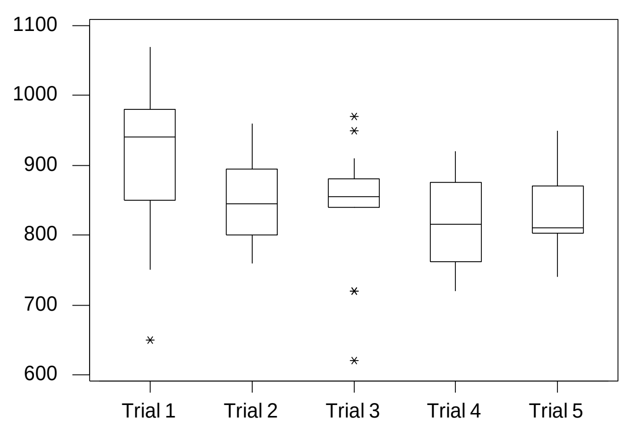
The data do not appear to be normally distributed. The points deviate from the line.

b)



After the transformation y* = log(y), the normal probability indicates that a normal distribution is reasonable.

6-121.

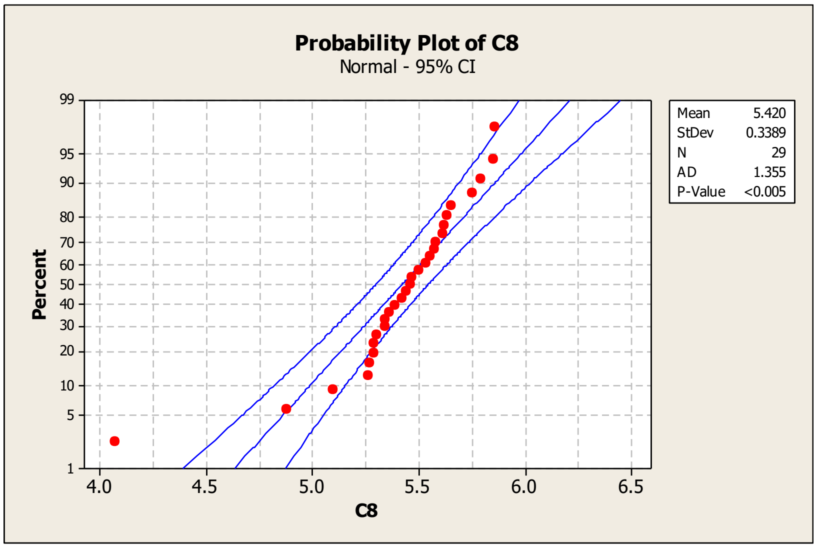


- There is a difference in the variability of the measurements in the trials. Trial 1 has the most variability in the measurements. Trial 3 has a small amount of variability in the main group of measurements, but there are four outliers. Trial 5 appears to have the least variability without any outliers.
- All of the trials except Trial 1 appear to be centered around 850. Trial 1 has a higher mean value
- All five trials appear to have measurements that are greater than the "true" value of 734.5.
- The difference in the measurements in Trial 1 may indicate a "start-up" effect in the data. There could be some bias in the measurements that is centering the data above the "true" value.

6-122. a) Descriptive Statistics

```
Variable        N  N*    Mean  SE Mean   StDev  Variance
Density        29   0  5.4197   0.0629  0.3389    0.1148
```

```
Variable      Minimum       Q1   Median       Q3  Maximum
Density        4.0700   5.2950   5.4600   5.6150   5.8600
```



b) There appears to be a low outlier in the data.

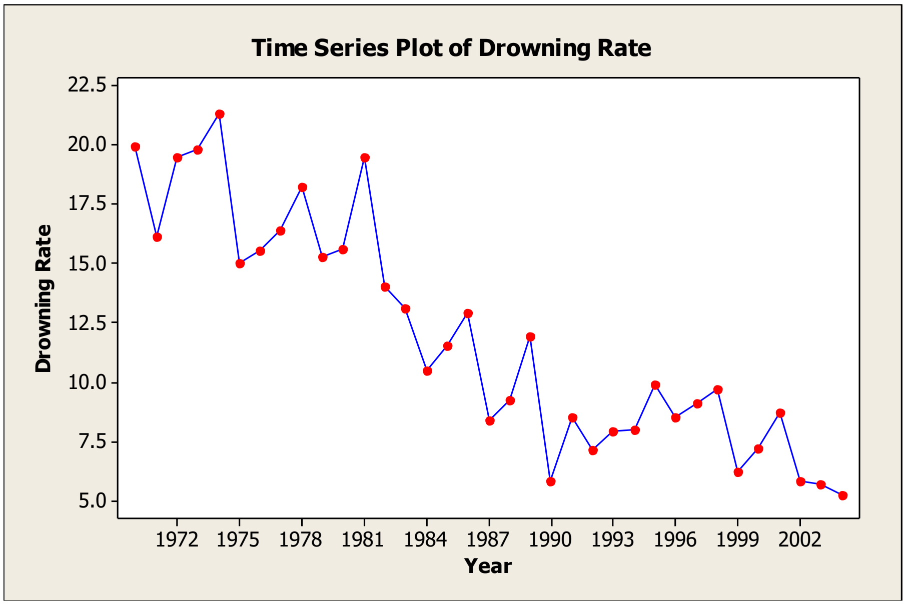
c) Due to the very low data point at 4.07, the mean may be lower than it should be. Therefore, the median would be a better estimate of the density of the earth. The median is not affected by a few outliers.

6-123.

a) Stem-and-leaf of Drowning Rate N = 35
Leaf Unit = 0.10

```
 4    5   2788
 5    6   2
 8    7   129
 13   8   04557
 17   9   1279
(1)   10  5
 17   11  59
 15   12  9
 14   13  1
 13   14  0
 12   15  0356
 8    16  14
 6    17
 6    18  2
 5    19  5589
 1    20
 1    21  3
```

Time Series Plots



b)

Descriptive Statistics: Drowning Rate

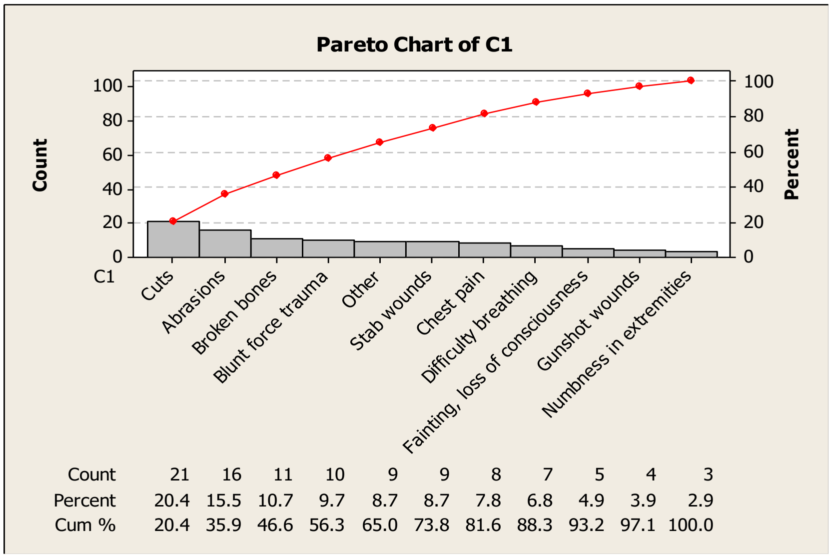
```
Variable        N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3
Drowning Rate  35   0  11.911    0.820  4.853    5.200   8.000  10.500  15.600

Variable       Maximum
Drowning Rate   21.300
```

c) Greater awareness of the dangers and drowning prevention programs might have been effective.

d) The summary statistics assume a stable distribution and may not adequately summarize the data because of the trend present.

6-124. a) Sort the categories by the number of instances in each category. Bars are used to indicate the counts and this sorted bar chart is known as a Pareto chart (discussed in Chapter 15).

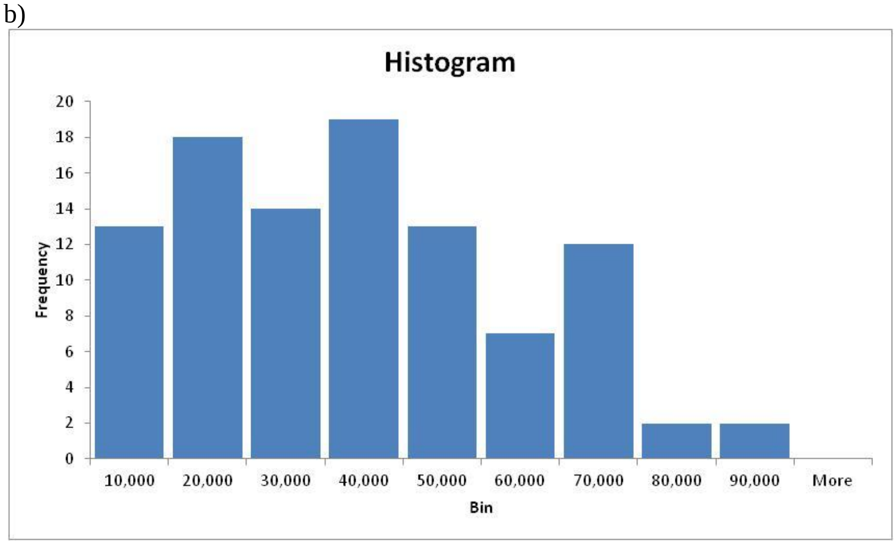
Cuts	21
Abrasions	16
Broken bones	11
Blunt force trauma	10
Other	9
Stab wounds	9
Chest pain	8
Difficulty breathing	7
Fainting, loss of consciousness	5
Gunshot wounds	4
Numbness in extremities	3



b) One would need to follow-up with patients that leave through a survey or phone calls to determine how long they waited before being seen and any other reasons that caused them to leave. This information could then be compiled and prioritized for improvements.

6-125. a) From computer software the sample mean = 34,232.05 and the sample standard deviation = 20,414.52

b)



There is substantial variability in mileage. There are number of vehicles with mileage near the mean, but another group with mileage near or even greater than 70,000.

c)
```
Stem-and-leaf of Mileage  N  = 100
Leaf Unit = 1000


 7     0  1223444
 13    0  567889
 16    1  013
 31    1  555555677889999
 35    2  1122
 45    2  5667888999
```

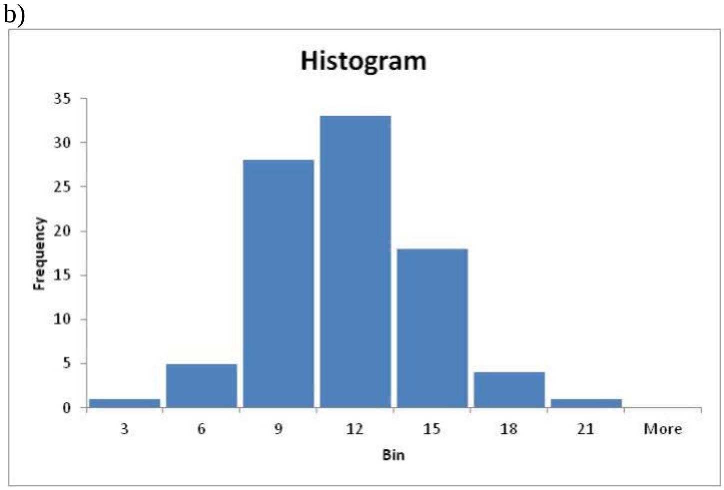
```
(13)   3   0001112223334
 42    3   667788
 36    4   0012334
 29    4   567899
 23    5   1114
 19    5   688
 16    6   00022334
 8     6   5677
 4     7   2
 3     7   7
 2     8   3
 1     8   5
```

d) For the given mileage of 45324, there are 29 observations greater and 71 observations less than this value. Therefore, the percentile is approximately 71%.

6-126.

a) The sample mean is 10.038 and the sample standard deviation is 2.868.

b)



The histogram is approximately symmetric around the mean value.

c)

```
Stem-and-leaf of Energy  N  = 90
Leaf Unit = 0.10


 1     2    9
 1     3
 2     4    0
 6     5    2599
 13    6    3677889
 23    7    1122666789
 34    8    02234556668
(13)   9    0234556788889
 43    10   022233344569
 31    11   01224677
 23    12   112366799
 14    13   1344469
 7     14   23
 5     15   12
 3     16   09
 1     17
 1     18   2
```

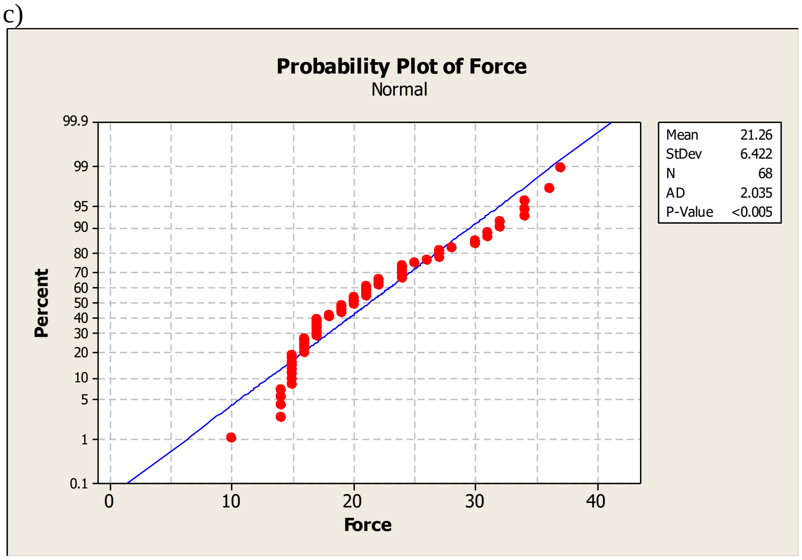
d) The mean plus two standard deviations equals 10.038 + 2(2.868) = 15.774. Here three data values exceed 15.774. Therefore, the proportion that exceeds 15.774 is 3/90 = 0.033.

6-127.

a)
```
Stem-and-leaf of Force   N  = 68
Leaf Unit = 1.0


 1    1  0
 1    1
 13   1  444455555555
 27   1  66666677777777
 33   1  889999
(9)   2  000011111
 26   2  222
 23   2  4444445
 16   2  6777
 12   2  8
 11   3  0011
 7    3  22
 5    3  444
 2    3  67
```

b) The sample mean is 21.265 and the sample standard deviation is 6.422.

c)

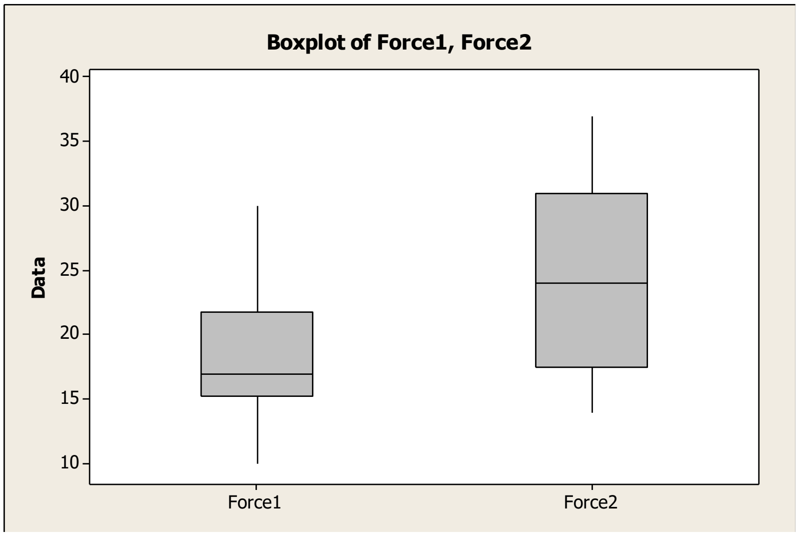


There a number of repeated values for force that are seen as points stacked vertically on the plot. This is probably due to round off of the force to two digits. There are fewer lower force values than are expected from a normal distribution.

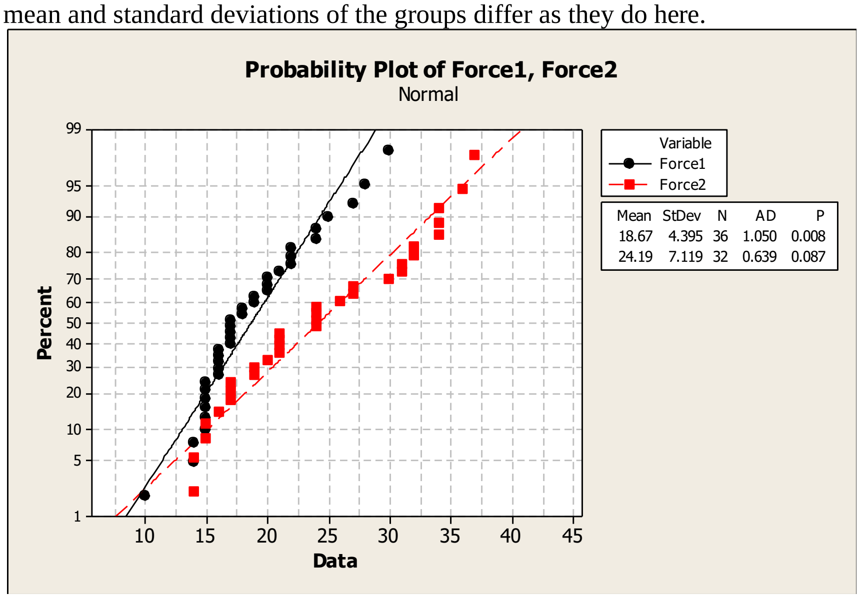
d) From the stem-and-leaf display, 9 caps exceed the force limit of 30. This is 9/68 = 0.132.

e) The mean plus two standard deviations equals 21.265 + 2(6.422) = 34.109. From the stem-and-leaf display, 2 caps exceed this force limit of 30. This is 2/68 = 0.029.

f) In the following box plots Force1 denotes the subset of the first 35 observations and Force2 denotes the remaining observation. The mean and variability of force is greater for the second set of data in Force2.

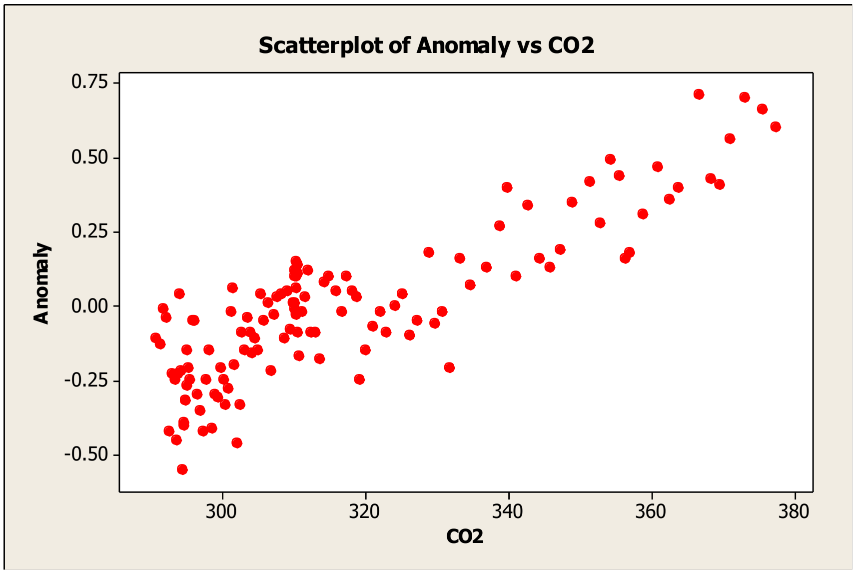


g) A separate normal distribution for each group of caps fits the data better. This is to be expected when the mean and standard deviations of the groups differ as they do here.



6-128.

a) The scatter plot indicates and approximately linear relationship.



b) The correlation coefficient is 0.852.

Mind Expanding Exercises

6-129. $\sum_{i=1}^{9} x_i^2 = 62572$ $\left(\sum_{i=1}^{9} x_i\right)^2 = 559504$ $n = 9$

$$s^2 = \frac{\sum_{i=1}^{9} x_i^2 - \frac{\left(\sum_{i=1}^{9} x_i\right)^2}{n}}{n-1} = \frac{62572 - \frac{559504}{9}}{9-1} = 50.61$$

$$s = \sqrt{50.61} = 7.11$$

Subtract 30 and multiply by 10

$\sum_{i=1}^{9} x_i^2 = 2579200$ $\left(\sum_{i=1}^{9} x_i\right)^2 = 22848400$ $n = 9$

$$s^2 = \frac{\sum_{i=1}^{9} x_i^2 - \frac{\left(\sum_{i=1}^{9} x_i\right)^2}{n}}{n-1} = \frac{2579200 - \frac{22848400}{9}}{9-1} = 5061.1$$

$$s = \sqrt{5061.1} = 71.14$$

Yes, the rescaling is by a factor of 10. Therefore, s^2 and s would be rescaled by multiplying s^2 by 10^2 (resulting in 5061.1) and s by 10 (71.14). Subtracting 30 from each value has no effect on the variance or standard deviation. This is because $V(aX + b) = a^2V(X)$.

6-130. $\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - a)^2$; The sum written in this form shows that the quantity is minimized when a = $\bar{x}$.

6-131. Of the two quantities $\sum_{i=1}^{n}(x_i - \bar{x})^2$ and $\sum_{i=1}^{n}(x_i - \mu)^2$, the quantity $\sum_{i=1}^{n}(x_i - \bar{x})^2$ will be smaller given that $\bar{x} \neq \mu$. This is because $\bar{x}$ is based on the values of the x_i's. The value of μ may be quite different for this sample.

6-132. $y_i = a + bx_i$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and} \quad \bar{y} = \frac{\sum_{i=1}^{n}(a + bx_i)}{n} = \frac{na + b\sum_{i=1}^{n} x_i}{n} = a + b\bar{x}$$

$$s_x^{\ 2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \text{ and } s_x = \sqrt{s_x^{\ 2}}$$

$$s_y^{\ 2} = \frac{\sum_{i=1}^{n}(a + bx_i - a - b\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(bx_i - b\bar{x})^2}{n-1} = \frac{b^2\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = b^2 s_x^{\ 2}$$

Therefore, $s_y = bs_x$

6-133. $\bar{x} = 835.00$ °F $s_x = 10.5$ °F

The results in °C:

$\bar{y} = -32 + 5/9\bar{x} = -32 + 5/9(835.00) = 431.89$ °C

$s_y^{\ 2} = b^2 s_x^{\ 2} = (5/9)^2(10.5)^2 = 34.028$ °C

6-134. Using the results found in a previous exercise with a = $-\frac{\bar{x}}{s}$ and b = 1/*s*, the mean and standard deviation of the z_i are $\bar{z} = 0$ and $s_Z = 1$.

6-135. Yes, in this case, since no upper bound on the last electronic component is available, use a measure of central location that is not dependent on this value. That measure is the median.

Sample Median = $\frac{x_{(4)} + x_{(5)}}{2} = \frac{63 + 75}{2} = 69$ *hours*

6-136. a) $\bar{x}_{n+1} = \frac{\sum_{i=1}^{n+1} x_i}{n+1} = \frac{\sum_{i=1}^{n} x_i + x_{n+1}}{n+1}$

$$\bar{x}_{n+1} = \frac{n\bar{x}_n + x_{n+1}}{n+1}$$

$$\bar{x}_{n+1} = \frac{n}{n+1}\bar{x}_n + \frac{x_{n+1}}{n+1}$$

b) $$ns_{n+1}^2 = \sum_{i=1}^{n} x_i^2 + x_{n+1}^2 - \frac{\left(\sum_{i=1}^{n} x_i + x_{n+1}\right)^2}{n+1}$$

$$= \sum_{i=1}^{n} x_i^2 + x_{n+1}^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n+1} - \frac{2x_{n+1}\sum_{i=1}^{n} x_i}{n+1} - \frac{x_{n+1}^2}{n+1}$$

$$= \sum_{i=1}^{n} x_i^2 + \frac{n}{n+1}x_{n+1}^2 - \frac{n}{n+1}2x_{n+1}\bar{x}_n - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n+1}$$

$$= \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n+1}\right] + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_{n+1}\bar{x}_n\right]$$

$$= \sum_{i=1}^{n} x_i^2 + \left[\frac{\left(\sum x_i\right)^2}{n} - \frac{\left(\sum x_i\right)^2}{n}\right] - \frac{\left(\sum x_i\right)^2}{n+1} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n} + \frac{(n+1)\left(\sum x_i\right)^2 - n\left(\sum x_i\right)^2}{n(n+1)} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= (n-1)s_n^2 + \frac{\left(\sum x_i\right)^2}{n(n+1)} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= (n-1)s_n^2 + \frac{n\bar{x}^2}{n+1} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= (n-1)s_n^2 + \frac{n}{n+1}\left(x^2{}_{n+1} - 2x_n\bar{x}_n + \bar{x}_n^2\right)$$

$$= (n-1)s_n^2 + \frac{n}{n+1}\left(x_{n+1}^2 - \bar{x}_n\right)^2$$

c) $\bar{x}_n = 65.811$ inches $\quad x_{n+1} = 64$

$$s_n^{\ 2} = 4.435 \qquad n = 37 \;\; s_n = 2.106$$

$$\bar{x}_{n+1} = \frac{37(65.81) + 64}{37 + 1} = 65.76$$

$$s_{n+1} = \sqrt{\frac{(37-1)4.435 + \frac{37}{37+1}(64 - 65.811)^2}{37}} = 2.098$$

6-137. The trimmed mean is pulled toward the median by eliminating outliers.
a) 10% Trimmed Mean = 89.29
b) 20% Trimmed Mean = 89.19
Difference is very small
c) No, the differences are very small, due to a very large data set with no significant outliers.

6-138.

If nT/100 is not an integer, calculate the two surrounding integer values and interpolate between the two. For example, if nT/100 = 2/3, one could calculate the mean after trimming 2 and 3 observations from each end and then interpolate between these two means.

CHAPTER 7

Section 7-2

7-1. The proportion of arrivals for chest pain is 8 among 103 total arrivals. The proportion = 8/103.

7-2. The proportion is 10/80 =1/8.

7-3. $$P(1.009 \le \overline{X} \le 1.012) = P\left(\tfrac{1.009-1.01}{0.003/\sqrt{9}} \le \tfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \tfrac{1.012-1.01}{0.003/\sqrt{9}}\right)$$
$$= P(-1 \le Z \le 2) = P(Z \le 2) - P(Z \le -1) = 0.9772 - 0.1586 = 0.8186$$

7-4. $X_i \sim N(100, 10^2)$ $\quad n = 25$

$$\mu_{\overline{X}} = 100 \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$
$$P[(100 - 1.8(2)) \le \overline{X} \le (100 + 2)] = P(96.4 \le \overline{X} \le 102) = P\left(\tfrac{96.4-100}{2} \le \tfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \tfrac{102-100}{2}\right)$$
$$= P(-1.8 \le Z \le 1) = P(Z \le 1) - P(Z \le -1.8) = 0.8413 - 0.0359 = 0.8054$$

7-5. $$\mu_{\overline{X}} = 75.5\,psi \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{6}} = 1.429$$
$$P(\overline{X} \ge 75.75) = P\left(\tfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \ge \tfrac{75.75-75.5}{1.429}\right)$$
$$= P(Z \ge 0.175) = 1 - P(Z \le 0.175)$$
$$= 1 - 0.56945 = 0.43055$$

7-6.

n = 6	n = 49
$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{6}} = 1.429$	$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{49}} = 0.5$
$\sigma_{\overline{X}}$ is reduced by 0.929 psi	

7-7. Assuming a normal distribution,

$$\mu_{\overline{X}} = 2500 \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{5}} = 22.361$$
$$P(2499 \le \overline{X} \le 2510) = P\left(\tfrac{2499-2500}{22.361} \le \tfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \tfrac{2510-2500}{22.361}\right)$$
$$= P(-0.045 \le Z \le 0.45) = P(Z \le 0.45) - P(Z \le -0.045)$$
$$= 0.6736 - 0.482 = 0.191$$

7-8. $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{5}} = 22.361\,psi$ = standard error of $\overline{X}$

7-9. $\sigma^2 = 25$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$n = \left(\frac{\sigma}{\sigma_{\overline{X}}}\right)^2 = \left(\frac{5}{1.5}\right)^2 = 11.11 \sim 12$$

7-10. Let $Y = \overline{X} - 6$

$$\mu_X = \frac{a+b}{2} = \frac{(0+1)}{2} = \tfrac{1}{2}$$

$$\mu_{\overline{X}} = \mu_X$$

$$\sigma_X^2 = \frac{(b-a)^2}{12} = \tfrac{1}{12}$$

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} = \frac{\frac{1}{12}}{12} = \frac{1}{144}$$

$$\sigma_{\overline{X}} = \tfrac{1}{12}$$

$$\mu_Y = \tfrac{1}{2} - 6 = -5\tfrac{1}{2}$$

$$\sigma_Y^2 = \tfrac{1}{144}$$

$Y = \overline{X} - 6 \sim N(-5\tfrac{1}{2}, \tfrac{1}{144})$, approximately, using the central limit theorem.

7-11. $n = 36$

$$\mu_X = \frac{a+b}{2} = \frac{(3+1)}{2} = 2$$

$$\sigma_X = \sqrt{\frac{1^2 + 0^2 + 1^2}{3}} = \sqrt{\frac{2}{3}}$$

$$\mu_{\overline{X}} = 2, \sigma_{\overline{X}} = \frac{\sqrt{2/3}}{\sqrt{36}} = \frac{\sqrt{2/3}}{6}$$

$$z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Using the central limit theorem:

$$P(2.1 < \overline{X} < 2.5) = P\left(\frac{2.1-2}{\frac{\sqrt{2/3}}{6}} < Z < \frac{2.5-2}{\frac{\sqrt{2/3}}{6}}\right) = P(0.7348 < Z < 3.6742)$$

$$= P(Z < 3.6742) - P(Z < 0.7348) = 1 - 0.7688 = 0.2312$$

7-12.

$\mu_X = 8.2$ minutes	n = 49
$\sigma_X = 1.5$ minutes	$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1.5}{\sqrt{49}} = 0.2143$ $\mu_{\overline{X}} = \mu_X = 8.2$ mins

Using the central limit theorem, $\overline{X}$ is approximately normally distributed.

a) $P(\overline{X} < 10) = P(Z < \frac{10 - 8.2}{0.2143}) = P(Z < 8.4) = 1$

b) $P(5 < \overline{X} < 10) = P(\frac{5-8.2}{0.2143} < Z < \frac{10-8.2}{0.2143})$

$= P(Z < 8.4) - P(Z < -14.932) = 1 - 0 = 1$

c) $P(\overline{X} < 6) = P(Z < \dfrac{6-8.2}{0.2143}) = P(Z < -10.27) = 0$

7-13.

$n_1 = 16$	$n_2 = 9$	$\overline{X}_1 - \overline{X}_2 \sim N(\mu_{\overline{X}_1} - \mu_{\overline{X}_2}, \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}) \sim N(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2})$
$\mu_1 = 75$	$\mu_2 = 70$	
$\sigma_1 = 8$	$\sigma_2 = 12$	$\sim N(75-70, \dfrac{8^2}{16} + \dfrac{12^2}{9}) \sim N(5, 20)$

a) $P(\overline{X}_1 - \overline{X}_2 > 4)$

$P(Z > \frac{4-5}{\sqrt{20}}) = P(Z > -0.2236) = 1 - P(Z \le -0.2236)$

$= 1 - 0.4115 = 0.5885$

b) $P(3.5 \le \overline{X}_1 - \overline{X}_2 \le 5.5)$

$P(\frac{3.5-5}{\sqrt{20}} \le Z \le \frac{5.5-5}{\sqrt{20}}) = P(Z \le 0.1118) - P(Z \le -0.3354)$

$= 0.5445 - 0.3687 = 0.1759$

7-14. If $\mu_B = \mu_A$, then $\overline{X}_B - \overline{X}_A$ is approximately normal with mean 0 and variance $\dfrac{\sigma_B^2}{25} + \dfrac{\sigma_A^2}{25} = 20.48$.

Then, $P(\overline{X}_B - \overline{X}_A > 3.5) = P(Z > \frac{3.5-0}{\sqrt{20.48}}) = P(Z > 0.773) = 0.2196$

The probability that $\overline{X}_B$ exceeds $\overline{X}_A$ by 3.5 or more is not that unusual when μ_B *and* μ_A are equal. Therefore, there is not strong evidence that μ_B is greater than μ_A.

7-15. Assume approximate normal distributions.

$(\overline{X}_{high} - \overline{X}_{low}) \sim N(60-55, \frac{4^2}{16} + \frac{4^2}{16})$

$\sim N(5, 2)$

$P(\overline{X}_{high} - \overline{X}_{low} \ge 2) = P(Z \ge \frac{2-5}{\sqrt{2}}) = 1 - P(Z \le -2.12) = 1 - 0.0170 = 0.983$

7-16. a) $SE_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1.60}{\sqrt{10}} = 0.51$

b) Here $\bar{X} \sim N(7.48, 0.51)$ and the standard normal distribution is used.

$$\begin{aligned} &P(6.49 < \bar{X} < 8.47) \\ &= P\left(\frac{6.49-7.48}{0.51} < \frac{\bar{X}-7.48}{0.51} < \frac{8.47-7.48}{0.51}\right) \\ &= P(-1.96 < Z < 1.96) = 0.975 - 0.025 = 0.95 \end{aligned}$$

c) We assume that $\bar{X}$ is normal distributed, or the sample size is sufficiently large that the central limit theorem applies.

7-17. The proportion of samples with pH below 5.0 . $\text{Proportion} = \frac{\text{Number of samples pH<5}}{\text{Number of total samples}} = \frac{26}{39} = 0.67$

7-18. a) $\bar{X} = 19.86, S_X = 23.65$, When $n = 8, SE_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{23.65}{\sqrt{8}} = 8.36$

b) Using the central limit theorem, $\bar{X} \sim N(19.86, 8.36)$
$P\big((19.86 - 8.36) < \bar{X} < (19.86 + 8.36)\big) = P(-1 < Z < 1) = 0.841 - 0.159 = 0.68$
where Z is a standard normal random variable.

c) The central limit theorem applies when the sample size n is large. Here n = 8 may be too small because the distribution of the counts of maple trees is quite skewed.

7-19. a) Point estimate of the mean proton flux is $\overline{\mathrm{X}} = 4958$
b) Point estimate of the standard deviation is $\mathrm{S_X} = 3420$
c) Estimate of the standard error is $\mathrm{S_{\overline{X}}} = \frac{\mathrm{S_X}}{\sqrt{25}} = \frac{3420}{5} = 684$
d) Point estimate of the median is 3360
e) Point estimate of the proportion of readings below 5000 is 16/25

7-20. a) Let $\overline{\mathrm{X}}$ denotes the mean miles and $\mathrm{E}(\overline{\mathrm{X}}) = \mu$, we further let Y denotes the additional miles
$\mathrm{E}(\overline{\mathrm{X}} + \mathrm{Y}) = \mathrm{E}(\overline{\mathrm{X}}) + \mathrm{E}(\mathrm{Y}) = \mu + 5 \cdot \mathrm{P(head)} + (-5) \cdot \mathrm{P(tail)}$
$= \mu + 5 \times 0.5 - 5 \times 0.5 = \mu + 0 = \mu$

b) Let Y denote the additional miles. The variance of Y is

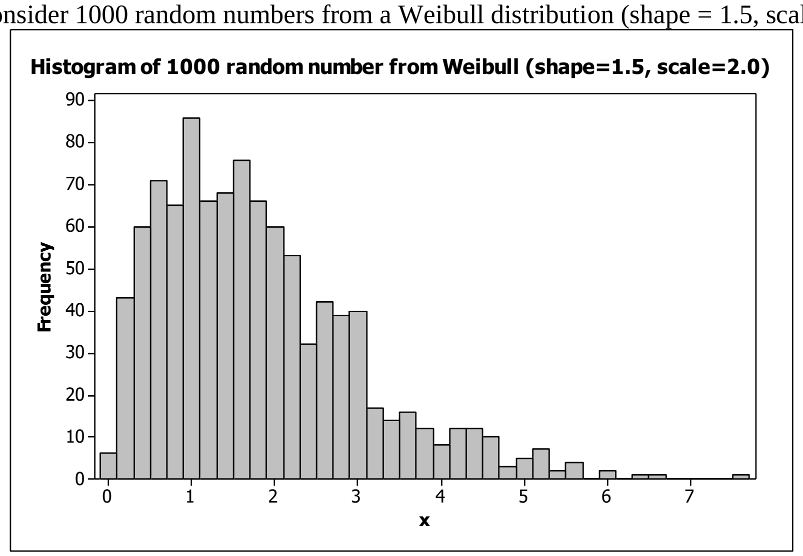
$$\sigma_{\mathrm{Y}}^2 = \mathrm{E}(\mathrm{Y}^2) - (\mathrm{EY})^2 = 5^2 \cdot \mathrm{P(head)} + (-5)^2 \cdot \mathrm{P(tail)} = \frac{25}{2} + \frac{25}{2} = 25$$

$\sigma_{\overline{\mathrm{X}}+\mathrm{Y}}^2 = \sigma_{\overline{\mathrm{X}}}^2 + \sigma_{\mathrm{Y}}^2 = \sigma_{\overline{\mathrm{X}}}^2 + 25.$ The standard deviation of Wayne's estimator is

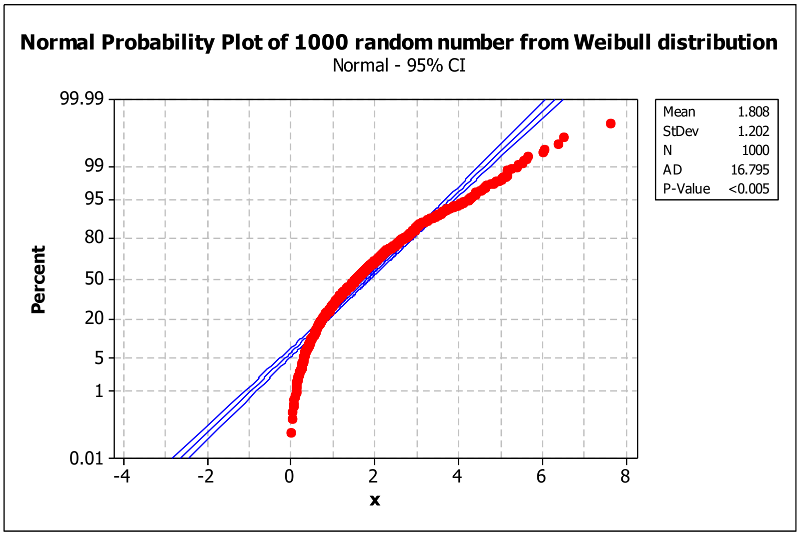
$\sqrt{\sigma_{\overline{\mathrm{X}}}^2 + 25} > \sigma_{\overline{\mathrm{X}}}$ and this is greater than the standard deviation of the sample mean.

c) Although Wayne's estimate is unbiased, it does not make good sense because it has a larger variance.

7-21. Consider 1000 random numbers from a Weibull distribution (shape = 1.5, scale = 2)



Normal probability plot

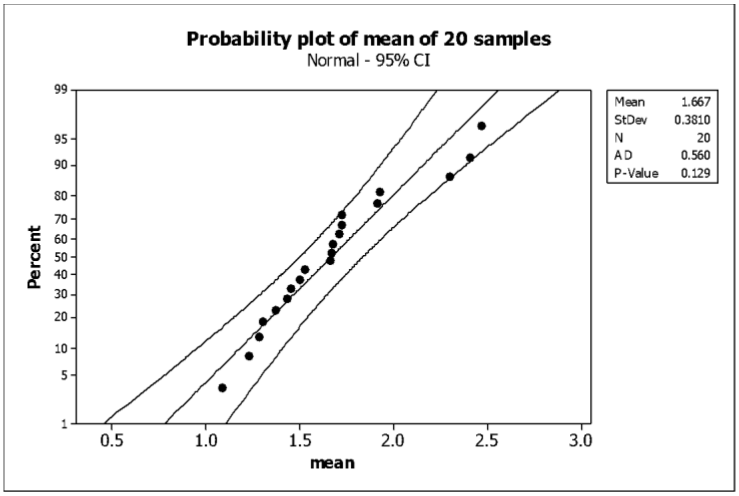


The data do not appear normally distributed. This Weibull distribution is skewed and has a long upper tail. Construct a table similar to Table 7-1

Obs	1	2	3	4	5	6	7	8	9	10
1	0.59	2.29	0.73	0.52	0.58	0.52	2.09	1.15	3.13	0.44
2	1.62	2.59	1.63	0.14	2.53	0.16	0.95	4.06	1.91	0.06
3	1.15	3.67	1.71	1.99	0.52	2.40	1.53	1.72	0.27	1.17
4	0.54	3.23	2.03	1.18	0.71	0.59	1.04	1.64	1.09	1.19
5	1.24	3.94	0.51	0.47	0.24	1.22	0.11	1.09	0.87	0.60
6	1.09	1.39	3.13	0.82	1.79	3.32	0.76	1.87	2.39	1.98
7	3.63	0.34	2.33	4.20	4.03	1.57	0.60	1.00	1.70	1.71
8	1.12	2.81	0.43	0.30	0.15	2.88	0.22	1.03	1.70	0.86
9	1.11	0.62	2.70	1.56	1.90	2.84	4.11	1.44	0.74	3.68
10	2.88	2.09	1.47	1.85	0.39	1.57	2.91	1.75	1.46	0.60
mean	1.50	2.30	1.67	1.30	1.28	1.71	1.43	1.68	1.53	1.23

Obs	11	12	13	14	15	16	17	18	19	20
1	3.19	1.27	2.03	0.97	1.92	3.28	1.68	1.75	3.08	0.10
2	0.82	1.12	0.39	3.78	3.69	1.22	1.45	1.39	3.97	0.80
3	3.55	0.78	0.79	2.05	3.89	1.51	0.34	3.57	2.03	0.66
4	3.36	0.06	1.80	0.62	1.18	0.73	0.67	0.42	1.99	3.04
5	0.47	4.51	3.61	2.02	1.71	0.71	4.08	3.76	2.77	0.12
6	0.27	1.23	0.94	3.68	3.08	0.70	0.09	1.78	2.53	2.72
7	1.85	1.89	1.77	0.38	3.05	2.35	1.00	2.27	1.34	0.48
8	0.70	1.93	1.16	0.40	3.07	1.24	3.86	0.78	3.04	0.71
9	4.16	2.10	0.20	2.98	0.48	1.66	2.44	1.51	0.57	0.69
10	0.89	1.73	1.01	0.36	2.57	1.15	1.64	1.91	2.71	1.54
mean	1.93	1.66	1.37	1.72	2.46	1.45	1.73	1.91	2.40	1.09

The normal probability plot of sample mean from each sample is much more normally distributed than the raw data.



Section 7-3

7-22. a) SE Mean $= \sigma_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{1.816}{\sqrt{20}} = 0.406$, Variance $= \sigma^2 = 1.816^2 = 3.298$

b) Estimate of mean of population = sample mean = 50.184

7-23. a) $\frac{S}{\sqrt{N}} = \text{SE Mean} \rightarrow \frac{10.25}{\sqrt{N}} = 2.05 \rightarrow N = 25$

$\text{Mean} = \frac{3761.70}{25} = 150.468$, $\text{Variance} = S^2 = 10.25^2 = 105.0625$

$\text{Variance} = \frac{\text{Sum of Squares}}{n-1} \rightarrow 105.0625 = \frac{SS}{25-1} \rightarrow SS = 2521.5$

b) Estimate of population mean = sample mean = 150.468

7-24. a) $E(\widehat{\Theta}_1) = \mathrm{E}\left(\frac{\mathrm{X}_1+\mathrm{X}_2}{2}\right) = \frac{1}{2}[\mathrm{E}(\mathrm{X}_1) + \mathrm{E}(\mathrm{X}_2)] = \frac{1}{2}[\mu + \mu] = \mu$

Therefore, $\widehat{\Theta}_1$is an unbiased estimator of μ

$E(\widehat{\Theta}_2) = \mathrm{E}\left(\frac{\mathrm{X}_1+3\mathrm{X}_2}{4}\right) = \frac{1}{4}[\mathrm{E}(\mathrm{X}_1) + 3\mathrm{E}(\mathrm{X}_2)] = \frac{1}{4}[\mu + 3\mu] = \mu$

Therefore $\widehat{\Theta}_2$is an unbiased estimator of μ

b) $V(\widehat{\Theta}_1) = \mathrm{V}\left(\frac{\mathrm{X}_1+\mathrm{X}_2}{2}\right) = \frac{1}{4}[\mathrm{V}(\mathrm{X}_1) + \mathrm{V}(\mathrm{X}_2)] = \frac{1}{4}[\sigma^2 + \sigma^2] = \frac{\sigma^2}{2}$

$$V(\widehat{\Theta}_2) = \mathrm{V}\left(\frac{\mathrm{X}_1 + 3\mathrm{X}_2}{4}\right) = \frac{1}{16}[\mathrm{V}(\mathrm{X}_1) + 3^2\mathrm{V}(\mathrm{X}_2)] = \frac{1}{16}[\sigma^2 + 9\sigma^2] = \frac{5\sigma^2}{8}$$

7-25. $E(\widehat{\Theta}) = E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2/c\right) = (n-1)\sigma^2/c$

$$\text{Bias} = E(\widehat{\Theta}) - \theta = \frac{(n-1)\sigma^2}{c} - \sigma^2 = \sigma^2\left(\frac{n-1}{c} - 1\right)$$

7-26. $$\mathrm{E}(\bar{X}_1) = E\left(\frac{\sum_{i=1}^{2n} X_i}{2n}\right) = \frac{1}{2n}E\left(\sum_{i=1}^{2n} X_i\right) = \frac{1}{2n}(2n\mu) = \mu$$

$$E\left(\overline{X}_2\right) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}(n\mu) = \mu$$

$\overline{X}_1$ and $\overline{X}_2$ are unbiased estimators of μ.

The variances are $V\left(\overline{X}_1\right) = \dfrac{\sigma^2}{2n}$ and $V\left(\overline{X}_2\right) = \dfrac{\sigma^2}{n}$; compare the MSE (variance in this case),

$$\frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} = \frac{\sigma^2/2n}{\sigma^2/n} = \frac{n}{2n} = \frac{1}{2}$$

Because both estimators are unbiased, one concludes that $\overline{X}_1$ is the "better" estimator with the smaller variance.

7-27. $E\left(\hat{\Theta}_1\right) = \frac{1}{7}\left[E(X_1) + E(X_2) + \cdots + E(X_7)\right] = \frac{1}{7}(7E(X)) = \frac{1}{7}(7\mu) = \mu$

$$E\left(\hat{\Theta}_2\right) = \frac{1}{2}\left[E(2X_1) + E(X_6) + E(X_7)\right] = \frac{1}{2}[2\mu - \mu + \mu] = \mu$$

a) Both $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimates of μ because the expected values of these statistics are equivalent to the true mean, μ.

b)

$$V\left(\hat{\Theta}_1\right) = V\left[\frac{X_1 + X_2 + ... + X_7}{7}\right] = \frac{1}{7^2}\left(V(X_1) + V(X_2) + \cdots + V(X_7)\right)$$

$$= \frac{1}{49}(7\sigma^2) = \frac{1}{7}\sigma^2$$

$$V(\hat{\Theta}_1) = \frac{\sigma^2}{7}$$

$$V\left(\hat{\Theta}_2\right) = V\left[\frac{2X_1 - X_6 + X_4}{2}\right] = \frac{1}{2^2}\left(V(2X_1) + V(X_6) + V(X_4)\right)$$

$$= \frac{1}{4}(4V(X_1) + V(X_6) + V(X_4))$$

$$= \frac{1}{4}\left(4\sigma^2 + \sigma^2 + \sigma^2\right) = \frac{1}{4}(6\sigma^2)$$

$$V(\hat{\Theta}_2) = \frac{3\sigma^2}{2}$$

Because both estimators are unbiased, the variances can be compared to select the better estimator. Because the variance of $\hat{\Theta}_1$ is smaller than that of $\hat{\Theta}_2$, $\hat{\Theta}_1$ is the better estimator.

7-28. Because both $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased, the variances of the estimators can compared to select the better estimator. Because the variance of $\hat{\theta}_2$ is smaller than that of $\hat{\theta}_1$, $\hat{\theta}_2$ is the better estimator.

Relative Efficiency = $\dfrac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} = \dfrac{V(\tilde{\Theta}_1)}{V(\hat{\Theta}_2)} = \dfrac{10}{4} = 2.5$

7-29. $E(\hat{\Theta}_1) = \theta \qquad E(\hat{\Theta}_2) = \theta/2$

$Bias = E(\hat{\Theta}_2) - \theta$

$= \dfrac{\theta}{2} - \theta = -\dfrac{\theta}{2}$

$V(\hat{\Theta}_1) = 10 \qquad V(\hat{\Theta}_2) = 4$

For unbiasedness, use $\hat{\Theta}_1$ because it is the only unbiased estimator.

As for minimum variance and efficiency we have

Relative Efficiency = $\dfrac{(V(\hat{\Theta}_1) + Bias^2)_1}{(V(\hat{\Theta}_2) + Bias^2)_2}$ where bias for θ_1 is 0.

Thus,

Relative Efficiency = $\dfrac{(10+0)}{\left(4 + \left(\dfrac{-\theta}{2}\right)^2\right)} = \dfrac{40}{\left(16+\theta^2\right)}$

If the relative efficiency is less than or equal to 1, $\hat{\Theta}_1$ is the better estimator.

Use $\hat{\Theta}_1$, when $\dfrac{40}{(16+\theta^2)} \le 1$

$$40 \le (16+\theta^2)$$
$$24 \le \theta^2$$
$$\theta \le -4.899 \text{ or } \theta \ge 4.899$$

If $-4.899 < \theta < 4.899$ then use $\hat{\Theta}_2$.

For unbiasedness, use $\hat{\Theta}_1$. For efficiency, use $\hat{\Theta}_1$ when $\theta \le -4.899$ or $\theta \ge 4.899$ and use $\hat{\Theta}_2$ when $-4.899 < \theta < 4.899$.

7-30.

$E(\hat{\Theta}_1) = \theta$	No bias	$V(\hat{\Theta}_1) = 12 = MSE(\hat{\Theta}_1)$
$E(\hat{\Theta}_2) = \theta$	No bias	$V(\hat{\Theta}_2) = 10 = MSE(\hat{\Theta}_2)$
$E(\hat{\Theta}_3) \ne \theta$	Bias	$MSE(\hat{\Theta}_3) = 6$ [note that this includes (bias^2)]

To compare the three estimators, calculate the relative efficiencies:

$\dfrac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} = \dfrac{12}{10} = 1.2$, because rel. eff. > 1 use $\hat{\Theta}_2$ as the estimator for θ

$\dfrac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_3)} = \dfrac{12}{6} = 2$, because rel. eff. > 1 use $\hat{\Theta}_3$ as the estimator for θ

$\dfrac{MSE(\hat{\Theta}_2)}{MSE(\hat{\Theta}_3)} = \dfrac{10}{6} = 1.8$, because rel. eff. > 1 use $\hat{\Theta}_3$ as the estimator for θ

Conclusion: $\hat{\Theta}_3$ is the most efficient estimator, but it is biased. $\hat{\Theta}_2$ is the best "unbiased" estimator.

7-31. $n_1 = 20$, $n_2 = 10$, $n_3 = 8$

Show that S^2 is unbiased.

$$E(S^2) = E\left(\frac{20S_1^2 + 10S_2^2 + 8S_3^2}{38}\right)$$
$$= \frac{1}{38}\left(E(20S_1^2) + E(10S_2^2) + E(8S_3^2)\right)$$
$$= \frac{1}{38}\left(20\sigma_1^2 + 10\sigma_2^2 + 8\sigma_3^2\right) = \frac{1}{38}\left(38\sigma^2\right) = \sigma^2$$

Therefore, S^2 is an unbiased estimator of σ^2 .

7-32. Show that $\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}$ is a biased estimator of σ^2

a)
$$E\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}\right)$$
$$= \frac{1}{n}E\left(\sum_{i=1}^{n}(X_i - n\overline{X})^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n}E(X_i^2) - nE(\overline{X}^2)\right) = \frac{1}{n}\left(\sum_{i=1}^{n}(\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right)$$
$$= \frac{1}{n}\left(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2\right) = \frac{1}{n}((n-1)\sigma^2) = \sigma^2 - \frac{\sigma^2}{n}$$

Therefore $\frac{\sum(X_i - \overline{X})^2}{n}$ is a biased estimator of σ^2

b) Bias $= E\left[\frac{\sum(X_i^2 - n\overline{X})^2}{n}\right] - \sigma^2 = \sigma^2 - \frac{\sigma^2}{n} - \sigma^2 = -\frac{\sigma^2}{n}$

c) Bias decreases as n increases.

7-33. a) Show that $\overline{X}^2$ is a biased estimator of μ^2. Using $\mathrm{E}(\mathrm{X}^2) = \mathrm{V(X)} + [\mathrm{E(X)}]^2$

$$E(\overline{X}^2) = \frac{1}{n^2}E\left(\sum_{i=1}^{n}X_i\right)^2 = \frac{1}{n^2}\left(V\left(\sum_{i=1}^{n}X_i\right) + \left[E\left(\sum_{i=1}^{n}X_i\right)\right]^2\right)$$
$$= \frac{1}{n^2}\left(n\sigma^2 + \left(\sum_{i=1}^{n}\mu\right)^2\right) = \frac{1}{n^2}\left(n\sigma^2 + (n\mu)^2\right)$$
$$= \frac{1}{n^2}\left(n\sigma^2 + n^2\mu^2\right)E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$$

Therefore, $\overline{X}^2$ is a biased estimator of μ.^{2}

b) Bias = $E(\overline{X}^2) - \mu^2 = \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \frac{\sigma^2}{n}$

c) Bias decreases as *n* increases.

7-34. a) The average of the 26 observations provided can be used as an estimator of the mean pull force because we know it is unbiased. This value is 75.615 pounds.

b) The median of the sample can be used as an estimate of the point that divides the population into a “weak” and “strong” half. This estimate is 75.2 pounds.

c) Our estimate of the population variance is the sample variance or 2.738 square pounds. Similarly, our estimate of the population standard deviation is the sample standard deviation or 1.655 pounds.

d) The estimated standard error of the mean pull force is $1.655/26^{1/2} = 0.325$. This value is the standard deviation, not of the pull force, but of the mean pull force of the sample.

e) Only one connector in the sample has a pull force measurement under 73 pounds. Our point estimate for the proportion requested is then 1/26 = 0.0385

7-35. **Descriptive Statistics**

```
Variable          N    Mean       Median      TrMean       StDev    SE Mean
Oxide Thickness 24   423.33      424.00      423.36        9.08       1.85
```

a) The mean oxide thickness, as estimated by Minitab from the sample, is 423.33 Angstroms.
b) The standard deviation for the population can be estimated by the sample standard deviation, or 9.08 Angstroms.
c) The standard error of the mean is 1.85 Angstroms.
d) Our estimate for the median is 424 Angstroms.
e) Seven of the measurements exceed 430 Angstroms, so our estimate of the proportion requested is 7/24 = 0.2917

7-36. a) $E(\hat{p}) = E(X/n) = \frac{1}{n}E(X) = \frac{1}{n}np = p$

b) The variance of $\hat{p}$ is $\frac{p(1-p)}{n}$ so its standard error must be $\sqrt{\frac{p(1-p)}{n}}$. To estimate this parameter we substitute our estimate of p into it.

7-37. a) $E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) - E(\overline{X}_2) = \mu_1 - \mu_2$

b) $s.e. = \sqrt{V(\overline{X}_1 - \overline{X}_2)} = \sqrt{V(\overline{X}_1) + V(\overline{X}_2) + 2COV(\overline{X}_1, \overline{X}_2)} = \sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}$

This standard error can be estimated by using the estimates for the standard deviations of populations 1 and 2.

c)

$$E(S_p^{\ 2}) = E\left(\frac{(n_1-1)\cdot S_1^{\ 2} + (n_2-1)\cdot S_2^{\ 2}}{n_1+n_2-2}\right) = \frac{1}{n_1+n_2-2}\left[(n_1-1)E(S_1^{\ 2}) + (n_2-1)\cdot E(S_2^{\ 2})\right] =$$

$$= \frac{1}{n_1+n_2-2}\left[(n_1-1)\cdot\sigma_1^{\ 2} + (n_2-1)\cdot\sigma_2^{\ 2})\right] = \frac{n_1+n_2-2}{n_1+n_2-2}\sigma^2 = \sigma^2$$

7-38. a) $E(\hat{\mu}) = E(\alpha\overline{X}_1 + (1-\alpha)\overline{X}_2) = \alpha E(\overline{X}_1) + (1-\alpha)E(\overline{X}_2) = \alpha\mu + (1-\alpha)\mu = \mu$

b)

$$s.e.(\hat{\mu}) = \sqrt{V(\alpha\overline{X}_1 + (1-\alpha)\overline{X}_2)} = \sqrt{\alpha^2 V(\overline{X}_1) + (1-\alpha)^2 V(\overline{X}_2)}$$
$$= \sqrt{\alpha^2 \frac{\sigma_1^{\ 2}}{n_1} + (1-\alpha)^2 \frac{\sigma_2^{\ 2}}{n_2}} = \sqrt{\alpha^2 \frac{\sigma_1^{\ 2}}{n_1} + (1-\alpha)^2 a \frac{\sigma_1^{\ 2}}{n_2}}$$
$$= \sigma_1 \sqrt{\frac{\alpha^2 n_2 + (1-\alpha)^2 a n_1}{n_1 n_2}}$$

c) The value of alpha that minimizes the standard error is $\alpha = \dfrac{an_1}{n_2 + an_1}$

d) With $a = 4$ and $n_1=2n_2$, the value of α to choose is 8/9. The arbitrary value of $\alpha = 0.5$ is too small and results in a larger standard error. With $\alpha = 8/9$, the standard error is

$$s.e.(\hat{\mu}) = \sigma_1 \sqrt{\frac{(8/9)^2 n_2 + (1/9)^2 8n_2}{2n_2^{\ 2}}} = \frac{0.667\sigma_1}{\sqrt{n_2}}$$

If $\alpha = 0.5$ the standard error is

$$s.e.(\hat{\mu}) = \sigma_1 \sqrt{\frac{(0.5)^2 n_2 + (0.5)^2 8n_2}{2n_2^{\ 2}}} = \frac{1.0607\sigma_1}{\sqrt{n_2}}$$

7-39.

a) $E(\frac{X_1}{n_1} - \frac{X_2}{n_2}) = \frac{1}{n_1}E(X_1) - \frac{1}{n_2}E(X_2) = \frac{1}{n_1}n_1 p_1 - \frac{1}{n_2}n_2 p_2 = p_1 - p_2 = E(p_1 - p_2)$

b) $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

c) An estimate of the standard error could be obtained substituting $\frac{X_1}{n_1}$ for p_1 and $\frac{X_2}{n_2}$ for p_2 in the equation shown in (b).

d) Our estimate of the difference in proportions is 0.01

e) The estimated standard error is 0.0413

7-40. $X \sim lognormal(1.5, 0.8^2)$, $n = 15, n_B = 200$, The original samples (n = 15):

#	1	2	3	4	5
Value	15.76	1.47	4.94	2.16	8.88
#	6	7	8	9	10
Value	1.32	13.40	4.14	4.10	4.19
#	11	12	13	14	15
Value	1.97	0.69	2.87	2.01	5.52

Here 200 bootstrap samples are generated and the sample median is computed for each. The standard error is estimated as the standard deviation of these sample medians, and in this case we obtain 1.04.

R code:

```
#generate the first 15 samples from the target distribution
n=15;
sample0=rlnorm(n, meanlog = 1.5, sdlog = 0.8);
#generate bootstrap samples and medians
nb=200
data=matrix(0, nb,n);
med=matrix(0, nb,1);
```

```
for (i in c(1: nb))
{
data[i,]=sample(sample0, n, replace = TRUE, prob = NULL)
med[i]=median(data[i,])
}
#calculate the standard error of the sample median "se"
se = sd(med)
```

7-41. $X \sim exp(\lambda = 0.1), n = 8, n_B = 100,$ the original sample (n = 8):

#	1	2	3	4	5	6	7	8
Value	1.88	4.27	18.15	8.37	26.25	5.76	0.74	7.45

Here 100 bootstrap samples are generated and the sample median is computed for each. The standard error is estimated as the standard deviation of these sample medians, and in this case we obtain 2.85.

R code:

```
#generate the first 8 samples from the target distribution
n=8
sample0=rexp(n, rate = 0.1)
#generate bootstrap samples and medians
nb=100
data=matrix(0,nb,n)
med=matrix(0,nb,1);
for (i in c(1:nb))
{
data[i,]=sample(sample0, n, replace = TRUE, prob = NULL)
med[i]=median(data[i,])
}
#calculate the standard error of the sample median "se"
se=sd(med)
```

7-42. $X \sim norm(\mu = 10,\ \sigma^2 = 4^2), n = 16,\ n_B = 200,$ the original sample (n = 16):

#	1	2	3	4	5	6	7	8
Value	4.26	6.59	12.36	7.47	10.84	1.17	17.02	14.10
#	9	10	11	12	13	14	15	16
Value	12.30	6.73	16.75	11.87	12.25	7.52	7.10	9.80

Here 200 bootstrap samples are generated and the sample mean is computed for each. The standard error is estimated as the standard deviation of these sample medians, and in this case we obtain 0.96. The bootstrap result is near 1, the true standard error.

R code

```
#generate the first 8 samples from the target distribution
n=16
nb=200
sample0=rnorm(n, mean = 10, sd=4)
#generate bootstrap samples and means
data=matrix(0,nb,n)
ave=matrix(0,nb,1);

for (i in c(1:nb))
{
```

```
data[i,]=sample(sample0, n, replace = TRUE, prob = NULL)
ave[i]=mean(data[i,])
}
#calculate the standard error of the sample mean
se=sd(ave)
```

7-43. Suppose that two independent random samples (of size n_1 and n_2) from two normal distributions are available. Explain how you would estimate the standard error of the difference in sample means $\bar{X}_1 - \bar{X}_2$ with the bootstrap method.

One case use the fact that $V(\bar{X}_1 - \bar{X}_2) = V(\bar{X}_1) + V(\bar{X}_2)$ (because the samples are independent) so that the standard error of $\bar{X}_1 - \bar{X}_2$ is $\sqrt{V(\bar{X}_1) + V(\bar{X}_2)}$. Then the bootstrap method can be used as in the previous exercise to estimate $V(\bar{X}_1)$ and $V(\bar{X}_2)$.

Section 7-4

7-44. $f(x) = p(1-p)^{x-1}$

$$L(p) = \prod_{i=1}^{n} p(1-p)^{x_i - 1} = p^n (1-p)^{\sum_{i=1}^{n} x_i - n}$$

$$\ln L(p) = n \ln p + \left(\sum_{i=1}^{n} x_i - n \right) \ln(1-p)$$

$$\frac{\partial \ln L(p)}{\partial p} = \frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p} \equiv 0$$

$$0 = \frac{(1-p)n - p\left(\sum_{i=1}^{n} x_i - n \right)}{p(1-p)} = \frac{n - np - p\sum_{i=1}^{n} x_i + pn}{p(1-p)}$$

$$0 = n - p\sum_{i=1}^{n} x_i$$

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$$

7-45. $f(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ $\qquad L(\lambda) = \prod_{i=1}^{n} \dfrac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \dfrac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$

$$\ln L(\lambda) = -n\lambda \ln e + \sum_{i=1}^{n} x_i \ln \lambda - \sum_{i=1}^{n} \ln x_i!$$

$$\frac{d \ln L(\lambda)}{d\lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i \equiv 0$$

$$= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\sum_{i=1}^{n} x_i = n\lambda$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

7-46. $f(x) = (\theta+1)x^{\theta}$

$$L(\theta) = \prod_{i=1}^{n} (\theta+1)x_i^{\theta} = (\theta+1)x_1^{\theta} \times (\theta+1)x_2^{\theta} \times \ldots = (\theta+1)^n \prod_{i=1}^{n} x_i^{\theta}$$

$$\ln L(\theta) = n\ln(\theta+1) + \theta \ln x_1 + \theta \ln x_2 + \ldots = n\ln(\theta+1) + \theta\sum_{i=1}^{n} \ln x_i$$

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{n}{\theta+1} + \sum_{i=1}^{n} \ln x_i = 0$$

$$\frac{n}{\theta+1} = -\sum_{i=1}^{n} \ln x_i$$

$$\hat{\theta} = \frac{n}{-\sum_{i=1}^{n} \ln x_i} - 1$$

7-47. $f(x) = \lambda e^{-\lambda(x-\theta)}$ for $x \ge \theta$ $\quad L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda(x-\theta)} = \lambda^n e^{-\lambda\sum_{i=1}^{n}(x-\theta)} = \lambda^n e^{-\lambda\left(\sum_{i=1}^{n} x - n\theta\right)}$

$$\ln L(\lambda,\theta) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i + \lambda n\theta$$

$$\frac{d\ln L(\lambda,\theta)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i + n\theta \equiv 0$$

$$\frac{n}{\lambda} = \sum_{i=1}^{n} x_i - n\theta$$

$$\hat{\lambda} = n\Big/\left(\sum_{i=1}^{n} x_i - n\theta\right)$$

$$\hat{\lambda} = \frac{1}{\bar{x} - \theta}$$

The other parameter θ cannot be estimated by setting the derivative of the log likelihood with respect to θ to zero because the log likelihood is a linear function of θ. The range of the likelihood is important.

The joint density function and therefore the likelihood is zero for θ < $Min(X_1, X_2, \ldots, X_n)$. The term in the log likelihood -*n*λθ is maximized for θ as small as possible within the range of nonzero likelihood. Therefore, the log likelihood is maximized for θ estimated with $Min(X_1, X_2, \ldots, X_n)$ so that $\hat{\theta} = x_{\min}$

b) Example: Consider traffic flow and let the time that has elapsed between one car passing a fixed point and the instant that the next car begins to pass that point be considered time headway. This headway can be modeled by the shifted exponential distribution.
Example in Reliability: Consider a process where failures are of interest. Suppose that a unit is put into operation at *x* = 0, but no failures will occur until θ time units of operation. Failures will occur only after the time θ.

7-48.

$$L(\theta) = \prod_{i=1}^{n} \frac{x_i e^{-x_i/\theta}}{\theta^2} \qquad \ln L(\theta) = \sum \ln(x_i) - \sum \frac{x_i}{\theta} - 2n \ln \theta$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{1}{\theta^2} \sum x_i - \frac{2n}{\theta}$$

Setting the last equation equal to zero and solving for theta yields

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{2n}$$

7-49. $E(X) = \dfrac{a-0}{2} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$, therefore: $\hat{a} = 2\overline{X}$

The expected value of this estimate is the true parameter, so it is unbiased. This estimate is reasonable in one sense because it is unbiased. However, there are obvious problems. Consider the sample $x_1=1$, $x_2 = 2$ and $x_3 = 10$. Now $\overline{x} = 4.37$ and $\hat{a} = 2\overline{x} = 8.667$. This is an unreasonable estimate of *a*, because clearly $a \geq 10$.

7-50. a) $\displaystyle\int_{-1}^{1} c(1+\theta x)dx = 1 = \left(cx + c\theta \frac{x^2}{2}\right)\Bigg|_{-1}^{1} = 2c$

so that the constant *c* should equal 0.5

b) $E(X) = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \dfrac{\theta}{3} \qquad \hat{\theta} = 3 \cdot \dfrac{1}{n}\sum_{i=1}^{n} X_i$

c) $E(\hat{\theta}) = E\left(3 \cdot \frac{1}{n}\sum_{i=1}^{n} X_i\right) = E(3\overline{X}) = 3E(\overline{X}) = 3\frac{\theta}{3} = \theta$

d)

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{2}(1+\theta X_i) \qquad \ln L(\theta) = n\ln(\frac{1}{2}) + \sum_{i=1}^{n}\ln(1+\theta X_i)$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{X_i}{(1+\theta X_i)}$$

By inspection, the value of θ that maximizes the likelihood is max (X_i)

7-51. a) $E(X^2) = 2\theta = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ so $\hat{\Theta} = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$

b)

$$L(\theta) = \prod_{i=1}^{n} \frac{x_i e^{-x_i^2/2\theta}}{\theta} \qquad \ln L(\theta) = \sum \ln(x_i) - \sum \frac{x_i^2}{2\theta} - n\ln\theta$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{1}{2\theta^2}\sum x_i^2 - \frac{n}{\theta}$$

Setting the last equation equal to zero, the maximum likelihood estimate is

$$\hat{\Theta} = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$$

and this is the same result obtained in part (a)

c)

$$\int_0^a f(x)dx = 0.5 = 1 - e^{-a^2/2\theta}$$

$$a = \sqrt{-2\theta\ln(0.5)} = \sqrt{2\theta\ln(2)}$$

We can estimate the median (a) by substituting our estimate for θ into the equation for a.

7-52. a) $\hat{a}$ cannot be unbiased since it will always be less than a.

b) bias = $\frac{na}{n+1} - \frac{a(n+1)}{n+1} = -\frac{a}{n+1} \underset{n\to\infty}{\to} 0$.

c) $2\overline{X}$

d) P(Y≤y)=P(X_1, …,X_n ≤ y)=[P(X_1≤y)]^n=$\left(\frac{y}{a}\right)^n$. Thus, f(y) is as given. Thus,

bias = E(Y) – a = $\frac{an}{n+1} - a = -\frac{a}{n+1}$.

e) For any $n > 1$, $n(n+2) > 3n$ so the variance of $\hat{a}_2$ is less than that of $\hat{a}_1$. It is in this sense that the second estimator is better than the first.

7-53. a)

$$L(\beta,\delta) = \prod_{i=1}^{n} \frac{\beta}{\delta}\left(\frac{x_i}{\delta}\right)^{\beta-1} e^{-\left(\frac{x_i}{\delta}\right)^{\beta}} = e^{-\sum_{i=1}^{n}\left(\frac{x_i}{\delta}\right)^{\beta}} \prod_{i=1}^{n} \frac{\beta}{\delta}\left(\frac{x_i}{\delta}\right)^{\beta-1}$$

$$\ln L(\beta,\delta) = \sum_{i=1}^{n} \ln\left[\frac{\beta}{\delta}\left(\frac{x_i}{\delta}\right)^{\beta-1}\right] - \sum\left(\frac{x_i}{\delta}\right)^{\beta} = n\ln\left(\frac{\beta}{\delta}\right) + (\beta-1)\sum \ln\left(\frac{x_i}{\delta}\right) - \sum\left(\frac{x_i}{\delta}\right)^{\beta}$$

b)

$$\frac{\partial \ln L(\beta,\delta)}{\partial \beta} = \frac{n}{\beta} + \sum \ln\left(\frac{x_i}{\delta}\right) - \sum \ln\left(\frac{x_i}{\delta}\right)\left(\frac{x_i}{\delta}\right)^{\beta}$$

$$\frac{\partial \ln L(\beta,\delta)}{\partial \delta} = -\frac{n}{\delta} - (\beta-1)\frac{n}{\delta} + \beta\frac{\sum x_i^{\beta}}{\delta^{\beta+1}}$$

Upon setting $\frac{\partial \ln L(\beta,\delta)}{\partial \delta}$ equal to zero, we obtain

$$\delta^{\beta} n = \sum x_i^{\beta} \quad \textit{and} \quad \delta = \left[\frac{\sum x_i^{\beta}}{n}\right]^{1/\beta}$$

Upon setting $\frac{\partial \ln L(\beta,\delta)}{\partial \beta}$ equal to zero and substituting for δ, we obtain

$$\frac{n}{\beta} + \sum \ln x_i - n\ln\delta = \frac{1}{\delta^{\beta}}\sum x_i^{\beta}(\ln x_i - \ln\delta)$$

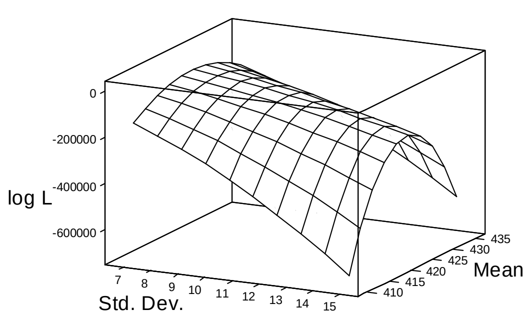
$$\frac{n}{\beta} + \sum \ln x_i - \frac{n}{\beta}\ln\left(\frac{\sum x_i^{\beta}}{n}\right) = \frac{n}{\sum x_i^{\beta}}\sum x_i^{\beta}\ln x_i - \frac{n}{\sum x_i^{\beta}}\sum x_i^{\beta}\frac{1}{\beta}\ln\left(\frac{\sum x_i^{\beta}}{n}\right)$$

and $\frac{1}{\beta} = \left[\frac{\sum x_i^{\beta}\ln x_i}{\sum x_i^{\beta}} + \frac{\sum \ln x_i}{n}\right]$

c) Numerical iteration is required.

7-54. a) Using the results from the example, we obtain that the estimate of the mean is 423.33 and the estimate of the variance is 82.4464

b)



The function has an approximate ridge and its curvature is not too pronounced. The maximum value for standard deviation is at 9.08, although it is difficult to see on the graph.

c) When *n* is increased to 40, the graph looks the same although the curvature is more pronounced. As *n* increases, it is easier to determine the maximum value for the standard deviation is on the graph.

7-55. From the example, the posterior distribution for μ is normal with mean $\dfrac{(\sigma^2/n)\mu_0+\sigma_0^2\bar{x}}{\sigma_0^2+\sigma^2/n}$ and variance $\dfrac{\sigma_0^2/(\sigma^2/n)}{\sigma_0^2+\sigma^2/n}$. The Bayes estimator for μ goes to the MLE as *n* increases. This follows because σ^2/n goes to 0, and the estimator approaches $\dfrac{\sigma_0^2\bar{x}}{\sigma_0^2}$ (the σ_0^2's cancel). Thus, in the limit $\hat{\mu}=\bar{x}$.

7-56. a) Because $f(x\mid\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ and $f(\mu)=\dfrac{1}{b-a}$ for $a\le\mu\le b$, the joint distribution is

$$f(x,\mu)=\frac{1}{(b-a)\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ for $-\infty<x<\infty$ and $a\le\mu\le b$.

Then, $$f(x)=\frac{1}{b-a}\int_a^b\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}d\mu$$

and this integral is recognized as a normal probability. Therefore,

$$f(x)=\frac{1}{b-a}\left[\Phi\left(\tfrac{b-x}{\sigma}\right)-\Phi\left(\tfrac{a-x}{\sigma}\right)\right]$$

where Φ(x) is the standard normal cumulative distribution function. Then

$$f(\mu\mid x)=\frac{f(x,\mu)}{f(x)}=\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma\left[\Phi\left(\frac{b-x}{\sigma}\right)-\Phi\left(\frac{a-x}{\sigma}\right)\right]}$$

b) The Bayes estimator is

$$\tilde{\mu}=\int_a^b\frac{\mu e^{-\frac{(x-\mu)^2}{2\sigma^2}}d\mu}{\sqrt{2\pi}\sigma\left[\Phi\left(\frac{b-x}{\sigma}\right)-\Phi\left(\frac{a-x}{\sigma}\right)\right]}.$$

Let v = (x - μ). Then, dv = - dμ and

$$\tilde{\mu}=\int_{x-b}^{x-a}\frac{(x-v)e^{-\frac{v^2}{2\sigma^2}}dv}{\sqrt{2\pi}\sigma\left[\Phi\left(\frac{b-x}{\sigma}\right)-\Phi\left(\frac{a-x}{\sigma}\right)\right]}=\frac{x\left[\Phi\left(\frac{x-a}{\sigma}\right)-\Phi\left(\frac{x-b}{\sigma}\right)\right]}{\left[\Phi\left(\frac{b-x}{\sigma}\right)-\Phi\left(\frac{a-x}{\sigma}\right)\right]}-\int_{x-b}^{x-a}\frac{ve^{-\frac{v^2}{2\sigma^2}}dv}{\sqrt{2\pi}\sigma\left[\Phi\left(\frac{b-x}{\sigma}\right)-\Phi\left(\frac{a-x}{\sigma}\right)\right]}$$

Let w = $\dfrac{v^2}{2\sigma^2}$. Then, $dw=[\frac{2v}{2\sigma^2}]dv=[\frac{v}{\sigma^2}]dv$ and

$$\tilde{\mu}=x-\int_{\frac{(x-b)^2}{2\sigma^2}}^{\frac{(x-a)^2}{2\sigma^2}}\frac{\sigma e^{-w}dw}{\sqrt{2\pi}\left[\Phi\left(\frac{b-x}{\sigma}\right)-\Phi\left(\frac{a-x}{\sigma}\right)\right]}=x+\frac{\sigma}{\sqrt{2\pi}}\left[\frac{e^{-\frac{(x-a)^2}{2\sigma^2}}-e^{-\frac{(x-b)^2}{2\sigma^2}}}{\Phi\left(\frac{b-x}{\sigma}\right)-\Phi\left(\frac{a-x}{\sigma}\right)}\right]$$

7-57. a) $f(x)=\dfrac{e^{-\lambda}\lambda^x}{x!}$ for x = 0, 1, 2, and $f(\lambda)=\left(\dfrac{m+1}{\lambda_0}\right)^{m+1}\dfrac{\lambda^m e^{-(m+1)\frac{\lambda}{\lambda_0}}}{\Gamma(m+1)}$ for λ > 0.

Then,

$$f(x,\lambda)=\frac{(m+1)^{m+1}\lambda^{m+x}e^{-\lambda-(m+1)\frac{\lambda}{\lambda_0}}}{\lambda_0^{m+1}\Gamma(m+1)x!}.$$

This last density is recognized to be a gamma density as a function of λ. Therefore, the posterior distribution of λ is a gamma distribution with parameters m + x + 1 and 1 + $\frac{m+1}{\lambda_0}$.

b) The mean of the posterior distribution can be obtained from the results for the gamma distribution to be

$$\frac{m+x+1}{\left[1+\frac{m+1}{\lambda_0}\right]}=\lambda_0\left(\frac{m+x+1}{m+\lambda_0+1}\right)$$

7-58 a) From the example, the Bayes estimate is $\tilde{\mu}=\dfrac{\frac{9}{25}(4)+1(4.85)}{\frac{9}{25}+1}=4.625$

b.) $\hat{\mu}=\bar{x}=4.85$ The Bayes estimate appears to underestimate the mean.

7-59. a) From the example, $\tilde{\mu}=\dfrac{(0.01)(5.03)+(\frac{1}{25})(5.05)}{0.01+\frac{1}{25}}=5.046$

b) $\hat{\mu}=\bar{x}=5.05$ The Bayes estimate is very close to the MLE of the mean.

7-60. a) $f(x\mid\lambda)=\lambda e^{-\lambda x},\ x\geq 0$ and $f(\lambda)=0.01e^{-0.01\lambda}$. Then,

$$f(x_1,x_2,\lambda)=\lambda^2e^{-\lambda(x_1+x_2)}0.01e^{-0.01\lambda}=0.01\lambda^2e^{-\lambda(x_1+x_2+0.01)}.$$

As a function of λ, this is recognized as a gamma density with parameters 3 and $x_1+x_2+0.01$. Therefore, the posterior mean for λ is

$$\tilde{\lambda}=\frac{3}{x_1+x_2+0.01}=\frac{3}{2\bar{x}+0.01}=0.00133.$$

b) Using the Bayes estimate for λ, P(X<1000)= $\int_0^{1000}0.00133e^{-.00133x}dx$ = 0.736

<u>Supplemental Exercises</u>

7-61. $f(x_1,x_2,...,x_n)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}\quad for\quad x_1>0,x_2>0,...,x_n>0$

7-62. $f(x_1,x_2,x_3,x_4,x_5)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^5\exp\left(-\sum_{i=1}^{5}\frac{(x_i-\mu)^2}{2\sigma^2}\right)$

7-63. $f(x_1, x_2, x_3, x_4) = 1 \quad for \quad 0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le x_3 \le 1, 0 \le x_4 \le 1$

7-64. $\bar{X}_1 - \bar{X}_2 \sim N(100-105, \frac{1.5^2}{25} + \frac{2^2}{30})$

$\sim N(-5, 0.2233)$

7-65. $X \sim N(50,144)$

$$P(47 \le \bar{X} \le 53) = P\left(\tfrac{47-50}{12/\sqrt{36}} \le Z \le \tfrac{53-50}{12/\sqrt{36}}\right) = P(-1.5 \le Z \le 1.5)$$

$$= P(Z \le 1.5) - P(Z \le -1.5) = 0.9332 - 0.0668 = 0.8664$$

No, because Central Limit Theorem states that with large samples (n ≥ 30), $\bar{X}$ is approximately normally distributed.

7-66. Assume $\bar{X}$ is approximately normally distributed.

$$P(\bar{X} > 4985) = 1 - P(\bar{X} \le 4985) = 1 - P(Z \le \tfrac{4985-5500}{100/\sqrt{9}})$$

$$= 1 - P(Z \le -15.45) = 1 - 0 = 1$$

7-67. $z = \dfrac{\bar{X} - \mu}{s/\sqrt{n}} = \dfrac{52-50}{\sqrt{2/16}} = 5.6569$

$P(Z > z) \approx 0$. The results are *very unusual*.

7-68. $P(\bar{X} \le 37) = P(Z \le -5.36) \approx 0$

7-69. Binomial with p equal to the proportion of defective chips and n = 100.

7-70. $E(a\bar{X}_1 + (1-a)\bar{X}_2 = a\mu + (1-a)\mu = \mu$

$$V(\bar{X}) = V[a\bar{X}_1 + (1-a)\bar{X}_2]$$

$$= a^2V(\bar{X}_1) + (1-a)^2V(\bar{X}_2) = a^2(\tfrac{\sigma^2}{n_1}) + (1-2a+a^2)(\tfrac{\sigma^2}{n_2})$$

$$= \frac{a^2\sigma^2}{n_1} + \frac{\sigma^2}{n_2} - \frac{2a\sigma^2}{n_2} + \frac{a^2\sigma^2}{n_2} = (n_2a^2 + n_1 - 2n_1a + n_1a^2)(\frac{\sigma^2}{n_1n_2})$$

$$\frac{\partial V(\bar{X})}{\partial a} = (\frac{\sigma^2}{n_1n_2})(2n_2a - 2n_1 + 2n_1a) \equiv 0$$

$$0 = 2n_2a - 2n_1 + 2n_1a$$

$$2a(n_2 + n_1) = 2n_1$$

$$a(n_2 + n_1) = n_1$$

$$a = \frac{n_1}{n_2 + n_1}$$

7-71.

$$L(\theta) = \left(\frac{1}{2\theta^3}\right)^n e^{\sum_{i=1}^{n} \frac{-x_i}{\theta}} \prod_{i=1}^{n} x_i^{\,2}$$

$$\ln L(\theta) = n \ln\left(\frac{1}{2\theta^3}\right) + 2\sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} \frac{x_i}{\theta}$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{-3n}{\theta} + \sum_{i=1}^{n} \frac{x_i}{\theta^2}$$

Making the last equation equal to zero and solving for θ, we obtain

$\hat{\Theta} = \dfrac{\sum_{i=1}^{n} x_i}{3n}$ as the maximum likelihood estimate.

7-72.

$$L(\theta) = \theta^n \prod_{i=1}^{n} x_i^{\theta-1}$$

$$\ln L(\theta) = n \ln \theta + (\theta - 1)\sum_{i=1}^{n} \ln(x_i)$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln(x_i)$$

making the last equation equal to zero and solving for theta, we obtain the maximum likelihood estimate

$$\hat{\Theta} = \frac{-n}{\sum_{i=1}^{n} \ln(x_i)}$$

7-73.

$$L(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} x_i^{\frac{1-\theta}{\theta}}$$

$$\ln L(\theta) = -n \ln \theta + \frac{1-\theta}{\theta} \sum_{i=1}^{n} \ln(x_i)$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = -\frac{n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^{n} \ln(x_i)$$

Upon setting the last equation equal to zero and solving for the parameter of interest, we obtain the maximum likelihood estimate

$$\hat{\Theta} = -\frac{1}{n} \sum_{i=1}^{n} \ln(x_i)$$

$$E(\hat{\theta}) = E\left[-\frac{1}{n}\sum_{i=1}^{n} \ln(x_i)\right] = \frac{1}{n} E\left[-\sum_{i=1}^{n} \ln(x_i)\right] = -\frac{1}{n}\sum_{i=1}^{n} E[\ln(x_i)]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \theta = \frac{n\theta}{n} = \theta$$

$E(\ln(X_i)) = \int_0^1 (\ln x) x^{\frac{1-\theta}{\theta}} dx$ let $u = \ln x$ and $dv = x^{\frac{1-\theta}{\theta}} dx$

then, $E(\ln(X)) = -\theta \int_0^1 x^{\frac{1-\theta}{\theta}} dx = -\theta$

7-74. a) Let $E\bar{X}^2 = \theta$. Then $V(\bar{X}) = E(\bar{X}^2) - (E\bar{X})^2$. Therefore $\sigma^2/n = \theta - \mu^2$ and $\theta = \sigma^2/n + \mu^2$

Therefore, $\bar{X}^2$ is a biased estimator of the area of the square.

b) $E(\bar{X}^2 - S^2/n) = \sigma^2/n + \mu^2 - E(S^2)/n = \mu^2$

7-75. $\hat{\mu} = \bar{x} = \frac{23.1+15.6+17.4+\cdots+28.7}{10} = 21.86$

Demand for all 5000 houses is $\theta = 5000\mu$
$\hat{\theta} = 5000\hat{\mu} = 5000(21.86) = 109{,}300$

The proportion estimate is $\hat{p} = \frac{7}{10} = 0.7$

Mind-Expanding Exercises

7-76. $P(X_1 = 0, X_2 = 0) = \dfrac{M(M-1)}{N(N-1)}$

$P(X_1 = 0, X_2 = 1) = \dfrac{M(N-M)}{N(N-1)}$

$P(X_1 = 1, X_2 = 0) = \dfrac{(N-M)M}{N(N-1)}$

$P(X_1 = 1, X_2 = 1) = \dfrac{(N-M)(N-M-1)}{N(N-1)}$

$P(X_1 = 0) = M/N$

$P(X_1 = 1) = \dfrac{N-M}{N}$

$$P(X_2 = 0) = P(X_2 = 0 \mid X_1 = 0)P(X_1 = 0) + P(X_2 = 0 \mid X_1 = 1)P(X_1 = 1)$$
$$= \frac{M-1}{N-1} \times \frac{M}{N} + \frac{M}{N-1} \times \frac{N-M}{N} = \frac{M}{N}$$
$$P(X_2 = 1) = P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0) + P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1)$$
$$= \frac{N-M}{N-1} \times \frac{M}{N} + \frac{N-M-1}{N-1} \times \frac{N-M}{N} = \frac{N-M}{N}$$

Because $P(X_2 = 0 \mid X_1 = 0) = \frac{M-1}{N-1}$ is not equal to $P(X_2 = 0) = \frac{M}{N}$, X_1 *and* X_2 are not independent.

7-77. a)

$$c_n = \frac{\Gamma[(n-1)/2]}{\Gamma(n/2)\sqrt{2/(n-1)}}$$

b) When $n = 10$, $c_n = 1.0281$. When $n = 25$, $c_n = 1.0105$. Therefore S is a reasonably good estimator for the standard deviation even when relatively small sample sizes are used.

7-78.

a) The likelihood is

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(x_i-\mu_i)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i-\mu_i)^2}{2\sigma^2}}$$

The log likelihood function is

$$-2\ln(L) = \sum_{i=1}^{n} [\frac{(x_i-\mu_i)^2}{\sigma^2} + \frac{(y_i-\mu_i)^2}{\sigma^2} + 4\ln\left(\sqrt{2\pi\sigma^2}\right)]$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}[(x_i-\mu_i)^2 + (y_i-\mu_i)^2] + 4n\ln\left(\sqrt{2\pi\sigma^2}\right)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}[x_i^2 + y_i^2 - 2\mu_i(x_i+y_i) + 2{\mu_i}^2] + 4n\ln\left(\sqrt{2\pi}\right) + 2n\ln\,(\sigma^2)$$

Take the derivative of each μ_i and set it to zero

$$\frac{\partial\,(-2\ln(L))}{\partial\mu_i} = \frac{-2(x_i+y_i)+4\mu_i}{\sigma^2} = 0$$

to obtain

$$\hat{\mu}_i = \frac{x_i+y_i}{2}$$

To find the maximum likelihood estimator of σ^2, substitute the estimate for μ_i and take the derivative with respect to σ^2

$$\frac{\partial\,(-2\ln(L))}{\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}[(x_i-\hat{\mu}_i)^2 + (y_i-\hat{\mu}_i)^2] + \frac{2n}{\sigma^2}$$

$$\frac{\partial\,(-2\ln(L))}{\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}\frac{2(x_i-y_i)^2}{4} + \frac{2n}{\sigma^2}$$

$$= -\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{2\sigma^4} + \frac{2n}{\sigma^2}$$

Set the derivative to zero and solve

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i-y_i)^2}{4n}$$

b)

$$E(\hat{\sigma}^2) = \frac{1}{4n}\sum_{i=1}^{n}E(y_i-x_i)^2 = \frac{1}{4n}\sum_{i=1}^{n}E(y_i^2 + x_i^2 - 2x_iy_i)$$

$$= \frac{1}{4n}\sum_{i=1}^{n}[E(y_i^2) + E(x_i^2) - E(2x_iy_i)] = \frac{1}{4n}\sum_{i=1}^{n}[\sigma^2 + \sigma^2 + 0] = \frac{\sigma^2}{2}$$

Therefore, the estimator is biased. The bias is independent of *n*.

c) An unbiased estimator of σ^2 is given by $2\hat{\sigma}^2$

7-79. $P\left(|\bar{X}-\mu| \ge \frac{c\sigma}{\sqrt{n}}\right) \le \frac{1}{c^2}$ from Chebyshev's inequality. Then, $P\left(|\bar{X}-\mu| < \frac{c\sigma}{\sqrt{n}}\right) \ge 1-\frac{1}{c^2}$. Given an ε, n and c can be chosen sufficiently large that the last probability is near 1 and $\frac{c\sigma}{\sqrt{n}}$ is equal to ε.

7-80. a) $P\left(X_{(n)} \le t\right) = P\left(X_i \le t \;\; for \;\; i=1,...,n\right) = [F(t)]^n$

$P\left(X_{(1)} > t\right) = P\left(X_i > t \;\; for \;\; i=1,...,n\right) = [1-F(t)]^n$

Then, $P\left(X_{(1)} \le t\right) = 1-[1-F(t)]^n$

b)

$$f_{X_{(1)}}(t) = \frac{\partial}{\partial t} F_{X_{(1)}}(t) = n[1-F(t)]^{n-1} f(t)$$

$$f_{X_{(n)}}(t) = \frac{\partial}{\partial t} F_{X_{(n)}}(t) = n[F(t)]^{n-1} f(t)$$

c) $P\left(X_{(1)} = 0\right) = F_{X_{(1)}}(0) = 1-[1-F(0)]^n = 1-p^n$ because F(0) = 1 - p.

$P\left(X_{(n)} = 1\right) = 1-F_{X_{(n)}}(0) = 1-[F(0)]^n = 1-(1-p)^n$

d) $P(X \le t) = F(t) = \Phi\left[\frac{t-\mu}{\sigma}\right]$. From a previous exercise,

$$f_{X_{(1)}}(t) = n\left\{1-\Phi\left[\frac{t-\mu}{\sigma}\right]\right\}^{n-1} \frac{1}{\sqrt{2\pi}\sigma} \; e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

$$f_{X_{(n)}}(t) = n\left\{\Phi\left[\frac{t-\mu}{\sigma}\right]\right\}^{n-1} \frac{1}{\sqrt{2\pi}\sigma} \; e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

e) $P(X \le t) = 1-e^{-\lambda t}$

From a previous exercise,

$F_{X_{(1)}}(t) = 1-e^{-n\lambda t}$	$f_{X_{(1)}}(t) = n\lambda e^{-n\lambda t}$
$F_{X_{(n)}}(t) = [1-e^{-\lambda t}]^n$	$f_{X_{(n)}}(t) = n[1-e^{-\lambda t}]^{n-1} \lambda e^{-\lambda t}$

7-81. $P\left(F\left(X_{(n)}\right) \le t\right) = P\left(X_{(n)} \le F^{-1}(t)\right) = t^n$ for $0 \le t \le 1$ from a previous exercise.

If $Y = F(X_{(n)})$, then $f_Y(y) = ny^{n-1}, 0 \le y \le 1$.

Then, $E(Y) = \int_0^1 ny^n dy = \frac{n}{n+1}$

$P\left(F\left(X_{(1)}\right) \le t\right) = P\left(X_{(1)} \le F^{-1}(t)\right) = 1-(1-t)^n \;\; 0 \le t \le 1$ from a previous exercise.

If $Y = F(X_{(1)})$, then $f_Y(y) = n(1-t)^{n-1}, 0 \le y \le 1$.

Then, $E(Y)=\int_0^1 yn(1-y)^{n-1}dy=\dfrac{1}{n+1}$ where integration by parts is used. Therefore,

$$E[F(X_{(n)})]=\frac{n}{n+1}\quad and \quad E[F(X_{(1)})]=\frac{1}{n+1}$$

7-82. $$E(V)=k\sum_{i=1}^{n-1}\ [E(X_{i+1}^2)+E(X_i^2)-2E(X_iX_{i+1})]$$

$$=k\sum_{i=1}^{n-1}\ (\sigma^2+\mu^2+\sigma^2+\mu^2-2\mu^2)=k(n-1)2\sigma^2$$

Therefore, $k=\frac{1}{2(n-1)}$

7-83. a) The traditional estimate of the standard deviation, *S*, is 3.26. The mean of the sample is 13.43 so the values of $\left|X_i-\overline{X}\right|$ corresponding to the given observations are 3.43, 1.43, 4.43, 0.57, 4.57, 1.57 and 2.57. The median of these new quantities is 2.57 so the new estimate of the standard deviation is 3.81 and this value is slightly larger than the value obtained from the traditional estimator.

b) Making the first observation in the original sample equal to 50 produces the following results. The traditional estimator, *S*, is equal to 13.91. The new estimator remains unchanged.

7-84. a)

$$\begin{aligned}T_r &= X_1 + \\ &X_1+X_2-X_1+ \\ &X_1+X_2-X_1+X_3-X_2+ \\ &\ldots+ \\ &X_1+X_2-X_1+X_3-X_2+\ldots+X_r-X_{r-1}+ \\ &(n-r)(X_1+X_2-X_1+X_3-X_2+\ldots+X_r-X_{r-1})\end{aligned}$$

Because X_1 is the minimum lifetime of *n* items, $E(X_1)=\dfrac{1}{n\lambda}$.

Then, $X_2 - X_1$ is the minimum lifetime of (*n*-1) items from the memoryless property of the exponential and $E(X_2-X_1)=\dfrac{1}{(n-1)\lambda}$.

Similarly, $E(X_k-X_{k-1})=\dfrac{1}{(n-k+1)\lambda}$. Then,

$$E(T_r)=\frac{n}{n\lambda}+\frac{n-1}{(n-1)\lambda}+\ldots+\frac{n-r+1}{(n-r+1)\lambda}=\frac{r}{\lambda}\text{ and }E\left(\frac{T_r}{r}\right)=\frac{1}{\lambda}=\mu$$

b) $V(T_r/r)=1/(\lambda^2 r)$ is related to the variance of the Erlang distribution

$V(X) = r / \lambda^2$. They are related by the value ($1/r^2$). The censored variance is ($1/r^2$) times the uncensored variance.

CHAPTER 8

Section 8-1

8-1

a) The confidence level for $\bar{x} - 2.14\sigma/\sqrt{n} \le \mu \le \bar{x} + 2.14\sigma/\sqrt{n}$ is determined by the value of z_0 which is 2.14. From Table III, Φ(2.14) = P(Z < 2.14) = 0.9838 and the confidence level is 2(0.9838-0.5) = 96.76%.

b) The confidence level for $\bar{x} - 2.49\sigma/\sqrt{n} \le \mu \le \bar{x} + 2.49\sigma/\sqrt{n}$ is determined by the value of z_0 which is 2.14. From Table III, Φ(2.49) = P(Z < 2.49) = 0.9936 and the confidence level is is 2(0.9936-0.5) = 98.72%.

c) The confidence level for $\bar{x} - 1.85\sigma/\sqrt{n} \le \mu \le \bar{x} + 1.85\sigma/\sqrt{n}$ is determined by the by the value of z_0 which is 2.14. From Table III, Φ(1.85) = P(Z < 1.85) = 0.9678 and the confidence level is 93.56%.

d) One-sided confidence interval with $z_\alpha = 2$. Therefore, α = P(Z > 2) = 0.0228 and confidence = 1 – α = = 0.9772 = 97.72%

e) One-sided confidence interval with $z_\alpha = 1.96$. Therefore, α = P(Z > 1.96) = 0.0250 and confidence = 1 – α = 0.9750 = 97.50%

8-2 a) A $z_\alpha = 2.33$ would result in a 98% two-sided confidence interval.
b) A $z_\alpha = 1.29$ would result in a 80% two-sided confidence interval.
c) A $z_\alpha = 1.15$ would result in a 75% two-sided confidence interval.

8-3 a) A $z_\alpha = 1.29$ would result in a 90% one-sided confidence interval.
b) A $z_\alpha = 1.65$ would result in a 95% one-sided confidence interval.
c) A $z_\alpha = 2.33$ would result in a 99% one-sided confidence interval.

8-4 a) 95% CI for μ, $n = 10, \quad \sigma = 20 \quad \bar{x} = 1000, \quad z = 1.96$

$$\bar{x} - z\sigma/\sqrt{n} \le \mu \le \bar{x} + z\sigma/\sqrt{n}$$

$$1000 - 1.96(20/\sqrt{10}) \le \mu \le 1000 + 1.96(20/\sqrt{10})$$

$$987.6 \le \mu \le 1012.4$$

b) .95% CI for μ, $n = 25, \quad \sigma = 20 \quad \bar{x} = 1000, \quad z = 1.96$

$$\bar{x} - z\sigma/\sqrt{n} \le \mu \le \bar{x} + z\sigma/\sqrt{n}$$

$$1000 - 1.96(20/\sqrt{25}) \le \mu \le 1000 + 1.96(20/\sqrt{25})$$

$$992.2 \le \mu \le 1007.8$$

c) 99% CI for μ, $n = 10, \quad \sigma = 20 \quad \bar{x} = 1000, \quad z = 2.58$

$$\bar{x} - z\sigma/\sqrt{n} \le \mu \le \bar{x} + z\sigma/\sqrt{n}$$

$$1000 - 2.58(20/\sqrt{10}) \le \mu \le 1000 + 2.58(20/\sqrt{10})$$

$$983.7 \le \mu \le 1016.3$$

d) 99% CI for μ, $n = 25, \quad \sigma = 20 \quad \bar{x} = 1000, \quad z = 2.58$

$$\bar{x} - z\sigma/\sqrt{n} \le \mu \le \bar{x} + z\sigma/\sqrt{n}$$
$$1000 - 2.58(20/\sqrt{25}) \le \mu \le 1000 + 2.58(20/\sqrt{25})$$
$$989.7 \le \mu \le 1010.3$$

e) When n is larger, the CI is narrower. The higher the confidence level, the wider the CI.

8-5 a) Sample mean from the first confidence interval = 38.02 + (61.98-38.02)/2 = 50
Sample mean from the second confidence interval = 39.95 + (60.05-39.95)/2 = 50

b) The 95% CI is (38.02, 61.98) and the 90% CI is (39.95, 60.05). The higher the confidence level, the wider the CI.

8-6 a) Sample mean from the first confidence interval =37.53 + (49.87-37.53)/2 = 43.7
Sample mean from the second confidence interval =35.59 + (51.81-35.59)/2 = 43.7

b) The 99% CI is (35.59, 51.81) and the 95% CI is (37.53, 49.87). The higher the confidence level, the wider the CI.

8-7 a) Find n for the length of the 95% CI to be 40. $Z_{a/2} = 1.96$

$1/2 \text{ length } = (1.96)(20)/\sqrt{n} = 20$

$39.2 = 20\sqrt{n}$

$n = \left(\frac{39.2}{20}\right)^2 = 3.84$

Therefore, $n = 4$.

b) Find n for the length of the 99% CI to be 40. $Z_{a/2} = 2.58$

$1/2 \text{ length } = (2.58)(20)/\sqrt{n} = 20$

$51.6 = 20\sqrt{n}$

$n = \left(\frac{51.6}{20}\right)^2 = 6.66$

Therefore, $n = 7$.

8-8 Interval (1): $3124.9 \le \mu \le 3215.7$ and Interval (2): $3110.5 \le \mu \le 3230.1$
Interval (1): half-length =90.8/2=45.4 and Interval (2): half-length =119.6/2=59.8

a) $\bar{x}_1 = 3124.9 + 45.4 = 3170.3$

$\bar{x}_2 = 3110.5 + 59.8 = 3170.3$ The sample means are the same.

b) Interval (1): $3124.9 \le \mu \le 3215.7$ was calculated with 95% confidence because it has a smaller half-length, and therefore a smaller confidence interval. The 99% confidence level widens the interval.

8-9 a) The 99% CI on the mean calcium concentration would be wider.

b) No, that is not the correct interpretation of a confidence interval. The probability that μ is between 0.49 and 0.82 is either 0 or 1.

c) Yes, this is the correct interpretation of a confidence interval. The upper and lower limits of the confidence limits are random variables.

8-10 95% Two-sided CI on the breaking strength of yarn: where $\bar{x}$ = 98, σ = 2 , n=9 and $z_{0.025}$ = 1.96

$$\bar{x} - z_{0.025}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{0.025}\sigma/\sqrt{n}$$
$$98 - 1.96(2)/\sqrt{9} \le \mu \le 98 + 1.96(2)/\sqrt{9}$$
$$96.7 \le \mu \le 99.3$$

8-11 95% Two-sided CI on the true mean yield: where $\bar{x}$ = 90.480, σ = 3 , n=5 and $z_{0.025}$ = 1.96

$$\bar{x} - z_{0.025}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{0.025}\sigma/\sqrt{n}$$
$$90.480 - 1.96(3)/\sqrt{5} \le \mu \le 90.480 + 1.96(3)/\sqrt{5}$$
$$87.85 \le \mu \le 93.11$$

8-12 99% Two-sided CI on the diameter cable harness holes: where $\bar{x}$ =1.5045 , σ = 0.01 , n=10 and $z_{0.005}$ = 2.58

$$\bar{x} - z_{0.005}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{0.005}\sigma/\sqrt{n}$$
$$1.5045 - 2.58(0.01)/\sqrt{10} \le \mu \le 1.5045 + 2.58(0.01)/\sqrt{10}$$
$$1.4963 \le \mu \le 1.5127$$

8-13 a) 99% Two-sided CI on the true mean piston ring diameter

For α = 0.01, $z_{\alpha/2} = z_{0.005}$ = 2.58 , and $\bar{x}$ = 74.036, σ = 0.001, n=15

$$\bar{x} - z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$74.036 - 2.58\left(\frac{0.001}{\sqrt{15}}\right) \le \mu \le 74.036 + 2.58\left(\frac{0.001}{\sqrt{15}}\right)$$
$$74.0353 \le \mu \le 74.0367$$

b) 99% One-sided CI on the true mean piston ring diameter

For α = 0.01, $z_{\alpha} = z_{0.01}$ =2.33 and $\bar{x}$ = 74.036, σ = 0.001, n=15

$$\bar{x} - z_{0.01}\frac{\sigma}{\sqrt{n}} \le \mu$$
$$74.036 - 2.33\left(\frac{0.001}{\sqrt{15}}\right) \le \mu$$
$$74.0354 \le \mu$$

The lower bound of the one-sided confidence interval is greater than the lower bound of the two-sided interval even though the level of significance is the same. This is because for a one-sided confidence interval the probability in the left tail (α) is greater than the probability in the left tail of the two-sided confidence interval (α/2).

8-14 a) 95% Two-sided CI on the true mean life of a 75-watt light bulb

For α = 0.05, $z_{\alpha/2} = z_{0.025}$ = 1.96, and $\bar{x}$ = 1014, σ =25, n=20

$$\bar{x} - z_{0.025}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{0.025}\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$1014 - 1.96\left(\frac{25}{\sqrt{20}}\right) \le \mu \le 1014 + 1.96\left(\frac{25}{\sqrt{20}}\right)$$
$$1003 \le \mu \le 1025$$

b) 95% one-sided CI on the true mean piston ring diameter

For $\alpha = 0.05$, $z_\alpha = z_{0.05} = 1.65$ and $\bar{x} = 1014$, $\sigma = 25$, $n=20$

$$\bar{x} - z_{0.05}\frac{\sigma}{\sqrt{n}} \le \mu$$
$$1014 - 1.65\left(\frac{25}{\sqrt{20}}\right) \le \mu$$
$$1005 \le \mu$$

The lower bound of the one-sided confidence interval is greater than the lower bound of the two-sided interval even though the level of significance is the same. This is because for a one-sided confidence interval the probability in the left tail (α) is greater than the probability in the left tail of the two-sided confidence interval ($\alpha/2$).

8-15 a) 95% two sided CI on the mean compressive strength
$z_{\alpha/2} = z_{0.025} = 1.96$, and $\bar{x} = 3250$, $\sigma^2 = 1000$, $n=12$

$$\bar{x} - z_{0.025}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{0.025}\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$3250 - 1.96\left(\frac{31.62}{\sqrt{12}}\right) \le \mu \le 3250 + 1.96\left(\frac{31.62}{\sqrt{12}}\right)$$
$$3232.11 \le \mu \le 3267.89$$

b) 99% Two-sided CI on the true mean compressive strength
$z_{\alpha/2} = z_{0.005} = 2.58$

$$\bar{x} - z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$3250 - 2.58\left(\frac{31.62}{\sqrt{12}}\right) \le \mu \le 3250 + 2.58\left(\frac{31.62}{\sqrt{12}}\right)$$
$$3226.4 \le \mu \le 3273.6$$

The 99% CI is wider than the 95% CI

8-16 95% Confident that the error of estimating the true mean life of a 75-watt light bulb is less than 5 hours.
For $\alpha = 0.05$, $z_{\alpha/2} = z_{0.025} = 1.96$, and $\bar{\sigma} = 25$, $E=5$

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96(25)}{5}\right)^2 = 96.04$$

Round up to the next integer. Therefore, $n = 97$

8-17 Set the width to 6 hours with $\sigma = 25$, $z_{0.025} = 1.96$ solve for n.

$$1/2 \text{ width } = (1.96)(25)/\sqrt{n} = 3$$
$$49 = 3\sqrt{n}$$
$$n = \left(\frac{49}{3}\right)^2 = 266.78$$

Therefore, n = 267

8-18 99% confidence that the error of estimating the true compressive strength is less than 15 psi
For $\alpha = 0.01$, $z_{\alpha/2} = z_{0.005} = 2.58$, and $\sigma = 31.62$, $E=15$

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{2.58(31.62)}{15}\right)^2 = 29.6 \cong 30$$

Therefore, $n = 30$

8-19 To decrease the length of the CI by one half, the sample size must be increased by 4 times (2^2).

$$z_{\alpha/2}\sigma/\sqrt{n} = 0.5l$$

Now, to decrease by half, divide both sides by 2.

$$(z_{\alpha/2}\sigma/\sqrt{n})/2 = (l/2)/2$$

$$(z_{\alpha/2}\sigma/2\sqrt{n}) = l/4$$

$$(z_{\alpha/2}\sigma/\sqrt{2^2 n}) = l/4$$

Therefore, the sample size must be increased by $2^2 = 4$

8-20 If n is doubled in Eq 8-7: $\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

$$\frac{z_{\alpha/2}\sigma}{\sqrt{2n}} = \frac{z_{\alpha/2}\sigma}{1.414\sqrt{n}} = \frac{z_{\alpha/2}\sigma}{1.414\sqrt{n}} = \frac{1}{1.414}\left(\frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right)$$

The interval is reduced by 0.293 or 29.3%

If n is increased by a factor of 4

$$\frac{z_{\alpha/2}\sigma}{\sqrt{4n}} = \frac{z_{\alpha/2}\sigma}{2\sqrt{n}} = \frac{z_{\alpha/2}\sigma}{2\sqrt{n}} = \frac{1}{2}\left(\frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right)$$

The interval is reduced by 0.5.

8-21 a) 99% two sided CI on the mean temperature
$z_{\alpha/2} = z_{0.005} = 2.57$, and $\bar{x} = 13.77$, $\sigma = 0.5$, $n=11$

$$\bar{x} - z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$13.77 - 2.57\left(\frac{0.5}{\sqrt{11}}\right) \le \mu \le 13.77 + 2.57\left(\frac{0.5}{\sqrt{11}}\right)$$

$$13.383 \le \mu \le 14.157$$

b) 95% lower-confidence bound on the mean temperature
For $\alpha = 0.05$, $z_{\alpha} = z_{0.05} = 1.65$ and $\bar{x} = 13.77$, $\sigma = 0.5$, $n = 11$

$$\bar{x} - z_{0.05}\frac{\sigma}{\sqrt{n}} \le \mu$$

$$13.77 - 1.65\left(\frac{0.5}{\sqrt{11}}\right) \le \mu$$

$$13.521 \le \mu$$

c) 95% confidence that the error of estimating the mean temperature for wheat grown is less than 2 degrees Celsius.
For $\alpha = 0.05$, $z_{\alpha/2} = z_{0.025} = 1.96$, and $\sigma = 0.5$, $E = 2$

$$n = \left(\frac{z_{a/2}\sigma}{E}\right)^2 = \left(\frac{1.96(0.5)}{2}\right)^2 = 0.2401$$

Round up to the next integer. Therefore n = 1.

d) Set the width to 1.5 degrees Celsius with σ = 0.5, $z_{0.025}$ = 1.96 solve for n.

$$1/2 \text{ width } = (1.96)(0.5)/\sqrt{n} = 0.75$$

$$0.98 = 0.75\sqrt{n}$$

$$n = \left(\frac{0.98}{0.75}\right)^2 = 1.707$$

Therefore, n = 2.

8-22 a) 95% CI for μ, $n = 5 \quad \sigma = 0.66 \quad \bar{x} = 18.56, \; z = 1.96$

$$\bar{x} - z\sigma/\sqrt{n} \le \mu \le \bar{x} + z\sigma/\sqrt{n}$$

$$3.372 - 1.96(0.66/\sqrt{5}) \le \mu \le 3.372 + 1.96(0.66/\sqrt{5})$$

$$2.79 \le \mu \le 3.95$$

b) Width is $2z\sigma/\sqrt{n} = 0.55$, therefore $n = [2z\sigma/0.55]2 = [2(1.96)(0.66)/0.55]2 = 22.13$
Round up to n = 23.

8-23 a) 99% CI for μ, $n = 12 \quad \sigma = 2.25 \quad \bar{x} = 28.0, \; z = 2.576$

$$\bar{x} - z\sigma/\sqrt{n} \le \mu \le \bar{x} + z\sigma/\sqrt{n}$$

$$28.0 - 2.576(2.25/\sqrt{12}) \le \mu \le 28.0 + 2.576(2.25/\sqrt{12})$$

$$26.33 \le \mu \le 29.67$$

b) Width is $2z\sigma/\sqrt{n} = 1.25$
Therefore z = $1.25n^{1/2}/(2\sigma) = 1.25(12^{1/2})/[2(2.25)] = 0.9623$
Therefore P(-0.9623 < Z < 0.9623) = 0.664 = 1 – α so that the confidence is 66.4%

Section 8-2

8-24 $t_{0.025,15} = 2.131$ $t_{0.05,10} = 1.812$ $t_{0.10,20} = 1.325$
$t_{0.005,25} = 2.787$ $t_{0.001,30} = 3.385$

8-25 a) $t_{0.025,12} = 2.179$ b) $t_{0.025,24} = 2.064$ c) $t_{0.005,13} = 3.012$
d) $t_{0.0005,15} = 4.073$

8-26 a) $t_{0.05,14} = 1.761$ b) $t_{0.01,19} = 2.539$ c) $t_{0.001,24} = 3.467$

8-27 a) Mean $= \frac{sum}{N} = \frac{251.848}{10} = 25.1848$
$Variance = (stDev)^2 = 1.605^2 = 2.5760$

b) 95% confidence interval on mean

$n = 10 \quad \bar{x} = 25.1848 \quad s = 1.605 \quad t_{0.025,9} = 2.262$

$$\bar{x} - t_{0.025,9}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,9}\left(\frac{s}{\sqrt{n}}\right)$$

$$25.1848 - 2.262\left(\frac{1.605}{\sqrt{10}}\right) \le \mu \le 25.1848 + 2.262\left(\frac{1.605}{\sqrt{10}}\right)$$

$$24.037 \le \mu \le 26.333$$

8-28 SE Mean $= \dfrac{stDev}{\sqrt{N}} = \dfrac{6.11}{\sqrt{N}} = 1.58$, therefore N = 15

Mean $= \dfrac{sum}{N} = \dfrac{751.40}{15} = 50.0933$

Variance $= (stDev)^2 = 6.11^2 = 37.3321$

b) 95% confidence interval on mean

$n = 15 \quad \bar{x} = 50.0933 \quad s = 6.11 \quad t_{0.025,14} = 2.145$

$$\bar{x} - t_{0.025,14}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,14}\left(\frac{s}{\sqrt{n}}\right)$$

$$50.0933 - 2.145\left(\frac{6.11}{\sqrt{15}}\right) \le \mu \le 50.0933 + 2.145\left(\frac{6.11}{\sqrt{15}}\right)$$

$$46.709 \le \mu \le 53.477$$

8-29 95% confidence interval on mean tire life

$n = 16 \quad \bar{x} = 60{,}139.7 \quad s = 3645.94 \quad t_{0.025,15} = 2.131$

$$\bar{x} - t_{0.025,15}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,15}\left(\frac{s}{\sqrt{n}}\right)$$

$$60139.7 - 2.131\left(\frac{3645.94}{\sqrt{16}}\right) \le \mu \le 60139.7 + 2.131\left(\frac{3645.94}{\sqrt{16}}\right)$$

$$58197.33 \le \mu \le 62082.07$$

8-30 99% lower confidence bound on mean Izod impact strength

$n = 20 \quad \bar{x} = 1.25 \quad s = 0.25 \quad t_{0.01,19} = 2.539$

$$\bar{x} - t_{0.01,19}\left(\frac{s}{\sqrt{n}}\right) \le \mu$$

$$1.25 - 2.539\left(\frac{0.25}{\sqrt{20}}\right) \le \mu$$

$$1.108 \le \mu$$

8-31 $\bar{x} = 1.10 \quad s = 0.015 \quad n = 25$

95% confidence interval on the mean volume of syrup dispensed

For $\alpha = 0.05$ and n = 25, $t_{\alpha/2,n-1} = t_{0.025,24} = 2.064$

$$\bar{x} - t_{0.025,24}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,24}\left(\frac{s}{\sqrt{n}}\right)$$
$$1.10 - 2.064\left(\frac{0.015}{\sqrt{25}}\right) \le \mu \le 1.10 + 2.064\left(\frac{0.015}{\sqrt{25}}\right)$$
$$1.094 \le \mu \le 1.106$$

8-32 95% confidence interval on mean peak power

$n = 7 \quad \bar{x} = 315 \quad s = 16 \quad t_{0.025,6} = 2.447$

$$\bar{x} - t_{0.025,6}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,6}\left(\frac{s}{\sqrt{n}}\right)$$
$$315 - 2.447\left(\frac{16}{\sqrt{7}}\right) \le \mu \le 315 + 2.447\left(\frac{315}{\sqrt{7}}\right)$$
$$300.202 \le \mu \le 329.798$$

8-33 99% upper confidence interval on mean SBP

$n = 14 \quad \bar{x} = 118.3 \quad s = 9.9 \quad t_{0.01,13} = 2.650$

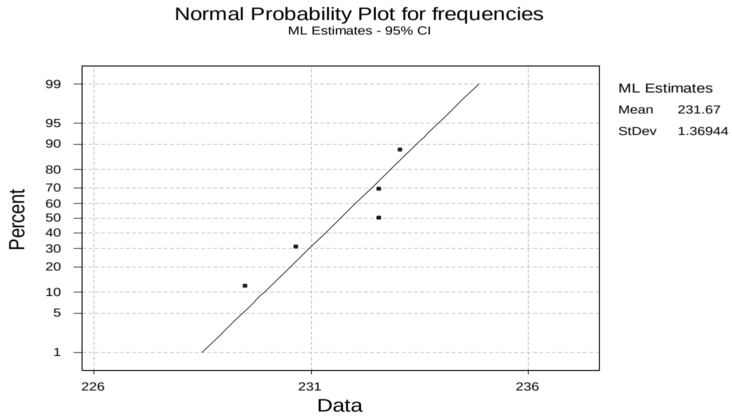
$$\mu \le \bar{x} + t_{0.005,13}\left(\frac{s}{\sqrt{n}}\right)$$
$$\mu \le 118.3 + 2.650\left(\frac{9.9}{\sqrt{14}}\right)$$
$$\mu \le 125.312$$

8-34 90% CI on the mean frequency of a beam subjected to loads

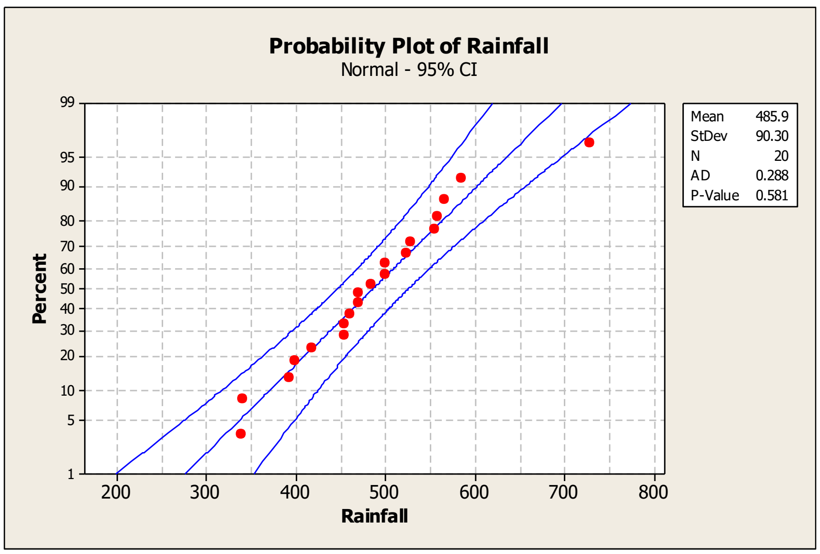
$\bar{x} = 231.67, \quad \text{s} = 1.53, \quad \text{n} = 5, \quad t_{\alpha/2,\text{n-1}} = t_{.05,4} = 2.132$

$$\bar{x} - t_{0.05,4}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.05,4}\left(\frac{s}{\sqrt{n}}\right)$$
$$231.67 - 2.132\left(\frac{1.53}{\sqrt{5}}\right) \le \mu \le 231.67 - 2.132\left(\frac{1.53}{\sqrt{5}}\right)$$
$$230.2 \le \mu \le 233.1$$

By examining the normal probability plot, it appears that the data are normally distributed.



8-35 The data appear to be normally distributed based on the normal probability plot below.



95% confidence interval on mean annual rainfall

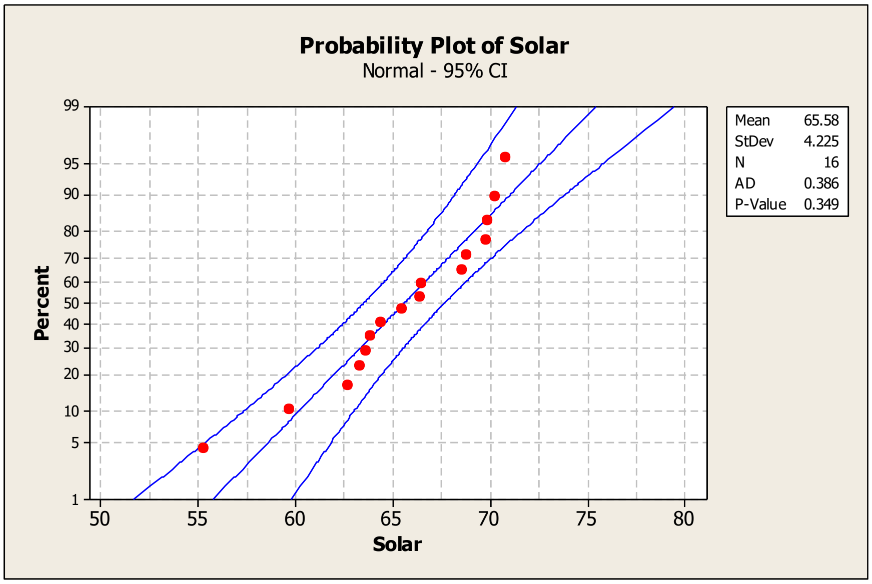
$$n = 20 \quad \bar{x} = 485.8 \quad s = 90.34 \quad t_{0.025,19} = 2.093$$

$$\bar{x} - t_{0.025,19}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,19}\left(\frac{s}{\sqrt{n}}\right)$$

$$485.8 - 2.093\left(\frac{90.34}{\sqrt{20}}\right) \le \mu \le 485.8 + 2.093\left(\frac{90.34}{\sqrt{20}}\right)$$

$$443.520 \le \mu \le 528.080$$

8-36 The data appear to be normally distributed based on the normal probability plot below.



95% confidence interval on mean solar energy consumed

$n = 16 \quad \bar{x} = 65.58 \quad s = 4.225 \quad t_{0.025,15} = 2.131$

$$\bar{x} - t_{0.025,15}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,15}\left(\frac{s}{\sqrt{n}}\right)$$

$$65.58 - 2.131\left(\frac{4.225}{\sqrt{16}}\right) \le \mu \le 65.58 + 2.131\left(\frac{4.225}{\sqrt{16}}\right)$$

$$63.329 \le \mu \le 67.831$$

8-37 99% confidence interval on mean current required

Assume that the data are a random sample from a normal distribution.

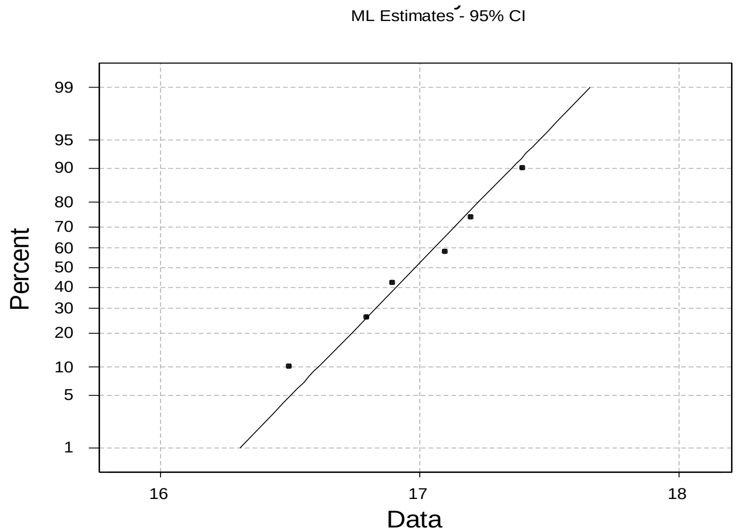
$n = 10 \quad \bar{x} = 317.2 \quad s = 15.7 \quad t_{0.005,9} = 3.250$

$$\bar{x} - t_{0.005,9}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.005,9}\left(\frac{s}{\sqrt{n}}\right)$$

$$317.2 - 3.250\left(\frac{15.7}{\sqrt{10}}\right) \le \mu \le 317.2 + 3.250\left(\frac{15.7}{\sqrt{10}}\right)$$

$$301.06 \le \mu \le 333.34$$

8-38 a) The data appear to be normally distributed based on the normal probability plot below.



b) 99% CI on the mean level of polyunsaturated fatty acid.
For $\alpha = 0.01$, $t_{\alpha/2,n-1} = t_{0.005,5} = 4.032$

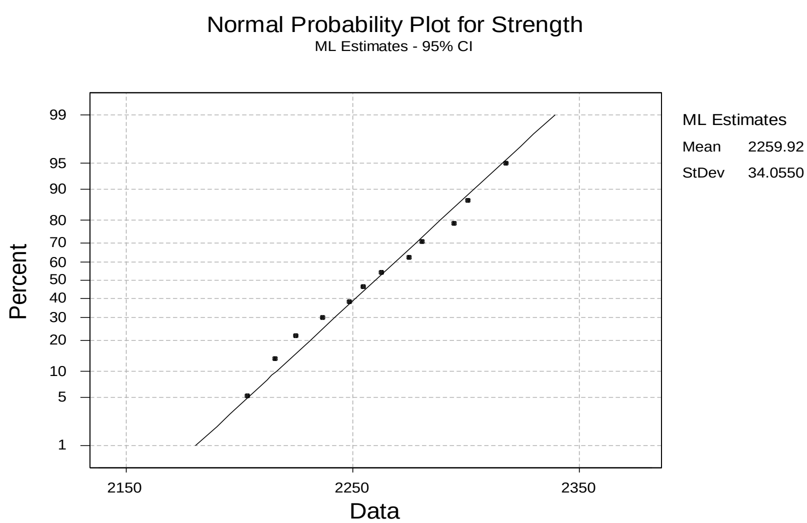
$$\bar{x} - t_{0.005,5}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.005,5}\left(\frac{s}{\sqrt{n}}\right)$$

$$16.98 - 4.032\left(\frac{0.319}{\sqrt{6}}\right) \le \mu \le 16.98 + 4.032\left(\frac{0.319}{\sqrt{6}}\right)$$

$$16.455 \le \mu \le 17.505$$

The 99% confidence for the mean polyunsaturated fat is (16.455, 17.505). There is high confidence that the true mean is in this interval

8-39 a) The data appear to be normally distributed based on examination of the normal probability plot below.



b) 95% two-sided confidence interval on mean comprehensive strength

$$n = 12 \quad \bar{x} = 2259.9 \quad s = 35.6 \quad t_{0.025,11} = 2.201$$

$$\bar{x} - t_{0.025,11}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,11}\left(\frac{s}{\sqrt{n}}\right)$$

$$2259.9 - 2.201\left(\frac{35.6}{\sqrt{12}}\right) \le \mu \le 2259.9 + 2.201\left(\frac{35.6}{\sqrt{12}}\right)$$

$$2237.3 \le \mu \le 2282.5$$

c) 95% lower-confidence bound on mean strength

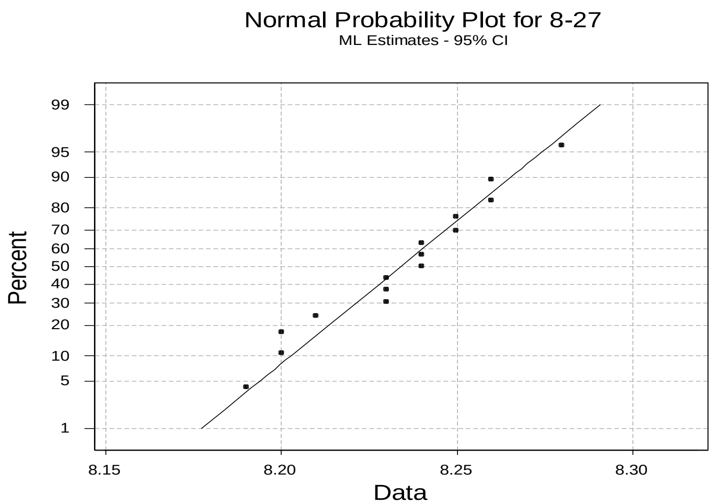
$$\bar{x} - t_{0.05,11}\left(\frac{s}{\sqrt{n}}\right) \le \mu$$

$$2259.9 - 1.796\left(\frac{35.6}{\sqrt{12}}\right) \le \mu$$

$$2241.4 \le \mu$$

8-40

a) According to the normal probability plot, there does not seem to be a severe deviation from normality for this data.



b) 95% two-sided confidence interval on mean rod diameter

For $\alpha = 0.05$ and n = 15, $t_{\alpha/2,n-1} = t_{0.025,14} = 2.145$

$$\bar{x} - t_{0.025,14}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,14}\left(\frac{s}{\sqrt{n}}\right)$$

$$8.23 - 2.145\left(\frac{0.025}{\sqrt{15}}\right) \le \mu \le 8.23 + 2.145\left(\frac{0.025}{\sqrt{15}}\right)$$

$$8.216 \le \mu \le 8.244$$

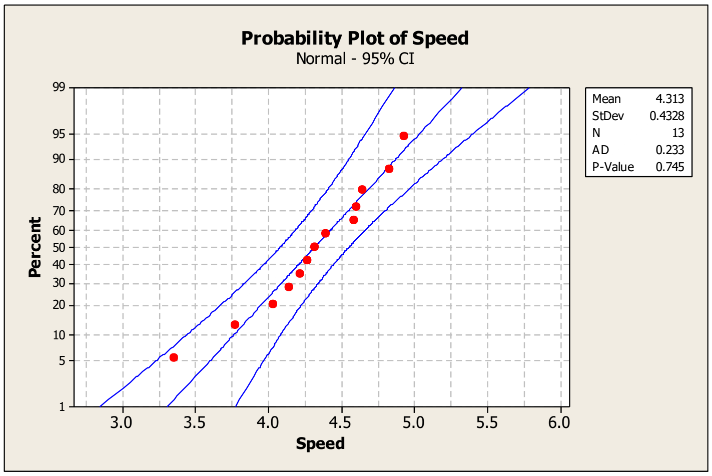
c) 95% upper confidence bound on mean rod diameter $t_{0.05,14} = 1.761$

$$\mu \le \bar{x} + t_{0.025,14}\left(\frac{s}{\sqrt{n}}\right)$$

$$\mu \le 8.23 + 1.761\left(\frac{0.025}{\sqrt{15}}\right)$$

$$\mu \le 8.241$$

8-41 a) The data appear to be normally distributed based on examination of the normal probability plot below.



b) 95% confidence interval on mean speed-up

$$n = 13 \quad \bar{x} = 4.313 \quad s = 0.4328 \quad t_{0.025,12} = 2.179$$

$$\bar{x} - t_{0.025,12}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,12}\left(\frac{s}{\sqrt{n}}\right)$$

$$4.313 - 2.179\left(\frac{0.4328}{\sqrt{13}}\right) \le \mu \le 4.313 + 2.179\left(\frac{0.4328}{\sqrt{13}}\right)$$

$$4.051 \le \mu \le 4.575$$

c) 95% lower confidence bound on mean speed-up

$$n = 13 \quad \bar{x} = 4.313 \quad s = 0.4328 \quad t_{0.05,12} = 1.782$$

$$\bar{x} - t_{0.05,12}\left(\frac{s}{\sqrt{n}}\right) \le \mu$$

$$4.313 - 1.782\left(\frac{0.4328}{\sqrt{13}}\right) \le \mu$$

$$4.099 \le \mu$$

8-42 95% lower bound confidence for the mean wall thickness given $\bar{x} = 4.05$, $s = 0.08$, $n = 25$

$t_{\alpha,n-1} = t_{0.05,24} = 1.711$

$$\bar{x} - t_{0.05,24}\left(\frac{s}{\sqrt{n}}\right) \le \mu$$

$$4.05 - 1.711\left(\frac{0.08}{\sqrt{25}}\right) \le \mu$$

$$4.023 \le \mu$$

There is high confidence that the true mean wall thickness is greater than 4.023 mm.

8-43 a) The data appear to be normally distributed.

b) 99% two-sided confidence interval on mean percentage enrichment

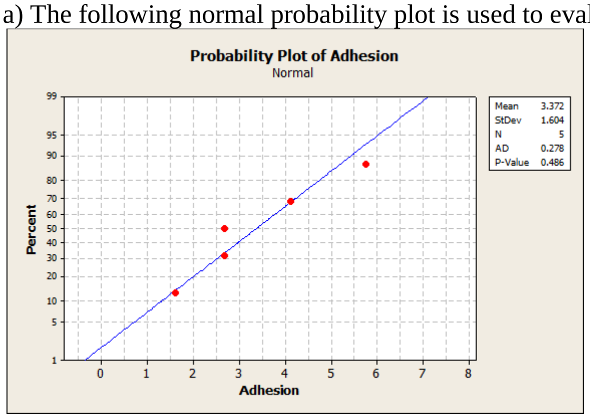
For $\alpha = 0.01$ and $n = 12$, $t_{\alpha/2,n-1} = t_{0.005,11} = 3.106$, $\bar{x} = 2.9017$ $s = 0.0993$

$$\bar{x} - t_{0.005,11}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.005,11}\left(\frac{s}{\sqrt{n}}\right)$$

$$2.902 - 3.106\left(\frac{0.0993}{\sqrt{12}}\right) \le \mu \le 2.902 + 3.106\left(\frac{0.0993}{\sqrt{12}}\right)$$

$$2.813 \le \mu \le 2.991$$

8-44 a) The following normal probability plot is used to evaluate the distribution.



There is no obvious deviation from normality.

b) 95% CI for μ, $n = 5$ $\bar{x} = 18.56$, $s = 1.604$, $t_{0.025,4} = 2.776$

$$\bar{x} - ts/\sqrt{n} \le \mu \le \bar{x} + ts/\sqrt{n}$$

$$3.372 - 2.776(1.604/\sqrt{5}) \le \mu \le 3.372 + 2.776(1.604/\sqrt{5})$$

$$1.38 \le \mu \le 5.36$$

8-45 a) 95% CI for $\mu, n = 12, \bar{x} = 2.082, s = 0.1564$, $t_{0.025,\,11} = 2.201$

$$\bar{x} - ts/\sqrt{n} \le \mu \le \bar{x} + ts/\sqrt{n}$$

$$2.082 - 2.210(0.1564/\sqrt{12}) \le \mu \le 2.082 + 2.210(0.1564/\sqrt{12})$$

$$1.98 \le \mu \le 2.18$$

b) The lower bound of 95% confidence interval is greater than the historical average of 1.95. Therefore, there is evidence that this clinic performs more CAT scans than usual.

A one-sided confidence interval would be more appropriate to answer this question. The one-sided interval follows.

$t_{0.05,\,11} = 1.7959$

$$\bar{x} - ts/\sqrt{n} \le \mu$$

$$2.082 - 1.7959(0.1564/\sqrt{12}) \le \mu$$

$$2.00 \le \mu$$

and the same conclusion is provided by this interval.

Section 8-3

8-46 $\chi^2_{0.05,10} = 18.31$ $\chi^2_{0.025,15} = 27.49$ $\chi^2_{0.01,12} = 26.22$

$\chi^2_{0.95,20} = 10.85$ $\chi^2_{0.99,18} = 7.01$ $\chi^2_{0.995,16} = 5.14$

$\chi^2_{0.005,25} = 46.93$

8-47 a) 95% upper CI and df = 24 $\chi^2_{1-\alpha,df} = \chi^2_{0.95,24} = 13.85$

b) 99% lower CI and df = 9 $\chi^2_{\alpha,df} = \chi^2_{0.01,9} = 21.67$

c) 90% CI and df = 19

$\chi^2_{\alpha/2,df} = \chi^2_{0.05,19} = 30.14$ and $\chi^2_{1-\alpha/2,df} = \chi^2_{0.95,19} = 10.12$

8-48 99% lower confidence bound for σ^2

For $\alpha = 0.01$ and n = 15, $\chi^2_{\alpha,n-1} = \chi^2_{0.01,14} = 29.14$

$$\frac{14(0.008)^2}{29.14} \le \sigma^2$$

$$0.00003075 \le \sigma^2$$

8-49 99% lower confidence bound for σ from the previous exercise is

$$0.00003075 \le \sigma^2$$

$$0.005545 \le \sigma$$

One may take the square root of the variance bound to obtain the confidence bound for the standard deviation.

8-50 95% two sided confidence interval for σ

$n = 10 \quad s = 4.8$

$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,9} = 19.02$ and $\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,9} = 2.70$

$$\frac{9(4.8)^2}{19.02} \le \sigma^2 \le \frac{9(4.8)^2}{2.70}$$

$$10.90 \le \sigma^2 \le 76.80$$

$$3.30 < \sigma < 8.76$$

8-51 95% confidence interval for σ given n = 51, s = 0.37

First find the confidence interval for σ^2

For $\alpha = 0.05$ and n = 51, $\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,50} = 71.42$ and $\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,50} = 32.36$

$$\frac{50(0.37)^2}{71.42} \le \sigma^2 \le \frac{50(0.37)^2}{32.36}$$

$0.096 \le \sigma^2 \le 0.2115$

Take the square root of the endpoints of this interval to obtain

$$0.31 < \sigma < 0.46$$

8-52 95% confidence interval for σ

$$n = 17 \quad s = 0.09$$

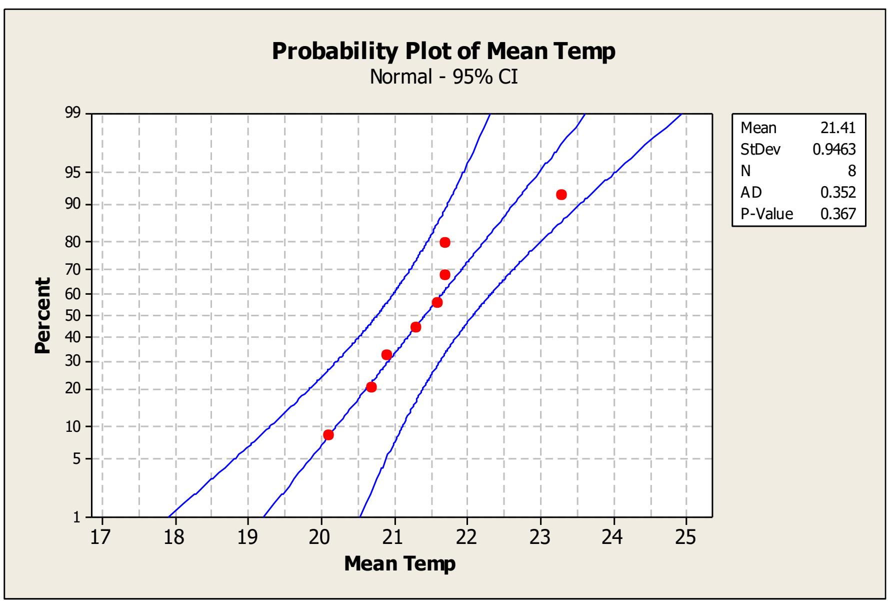
$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,16} = 28.85 \text{ and } \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,16} = 6.91$$

$$\frac{16(0.09)^2}{28.85} \le \sigma^2 \le \frac{16(0.09)^2}{6.91}$$

$$0.0045 \le \sigma^2 \le 0.0188$$

$$0.067 < \sigma < 0.137$$

8-53 The data appear to be normally distributed based on examination of the normal probability plot below.



95% confidence interval for σ

$$n = 8 \quad s = 0.9463$$

$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,7} = 16.01 \text{ and } \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,7} = 1.69$$

$$\frac{7(0.9463)^2}{16.01} \le \sigma^2 \le \frac{7(0.9463)^2}{1.69}$$

$$0.392 \le \sigma^2 \le 3.709$$

$$0.626 < \sigma < 1.926$$

8-54 95% confidence interval for σ

$$n = 41 \quad s = 15.99$$

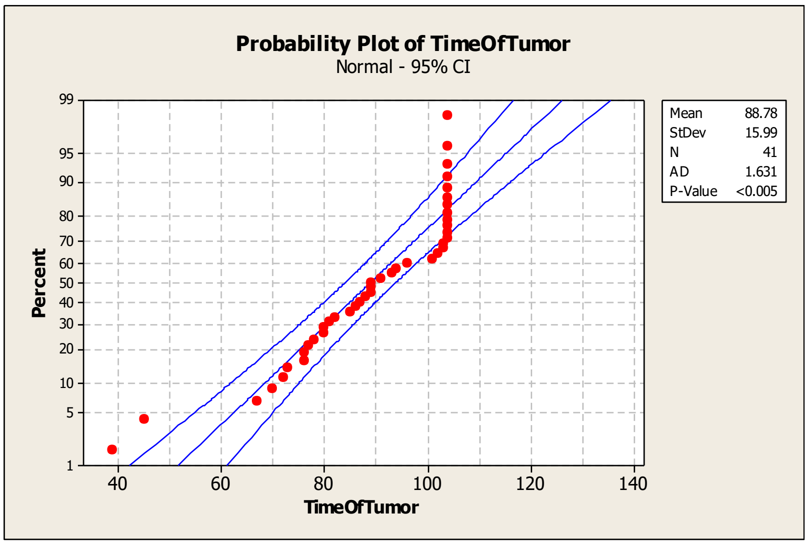
$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,40} = 59.34 \text{ and } \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,40} = 24.43$$

$$\frac{40(15.99)^2}{59.34} \le \sigma^2 \le \frac{40(15.99)^2}{24.43}$$

$$172.35 \le \sigma^2 \le 418.633$$

$$13.13 < \sigma < 20.46$$

The data do not appear to be normally distributed based on examination of the normal probability plot below. Therefore, the 95% confidence interval for σ is invalid.



8-55 95% confidence interval for σ

$$n = 15 \quad s = 0.00831$$

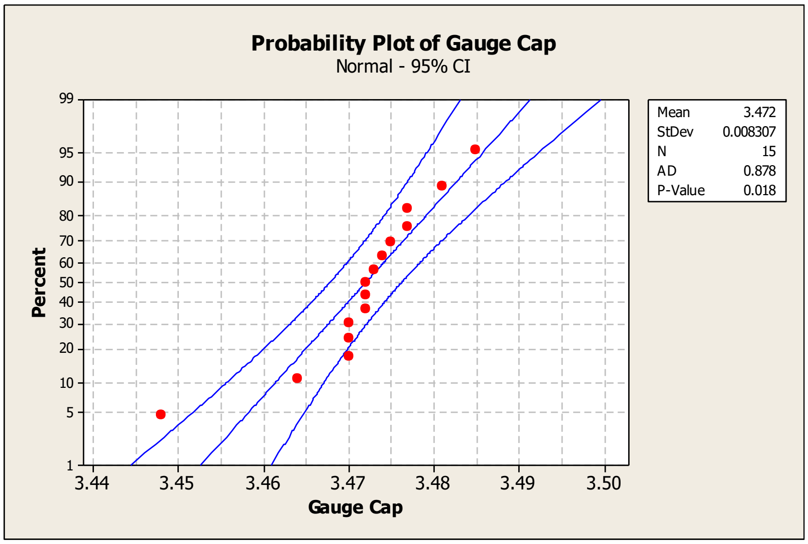
$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,14} = 26.12 \text{ and } \chi^2_{1-\alpha,n-1} = \chi^2_{0.95,14} = 6.53$$

$$\sigma^2 \le \frac{14(0.00831)^2}{6.53}$$

$$\sigma^2 \le 0.000148$$

$$\sigma \le 0.0122$$

The data do not appear to be normally distributed based on an examination of the normal probability plot below. Therefore, the 95% confidence interval for σ is not valid.



8-56 a) 99% two-sided confidence interval on σ^2

$$n = 10 \qquad s = 1.913 \quad \chi^2_{0.005,9} = 23.59 \text{ and } \chi^2_{0.995,9} = 1.73$$

$$\frac{9(1.913)^2}{23.59} \le \sigma^2 \le \frac{9(1.913)^2}{1.73}$$

$$1.396 \le \sigma^2 \le 19.038$$

b) 99% lower confidence bound for σ^2

For $\alpha = 0.01$ and n = 10, $\chi^2_{\alpha,n-1} = \chi^2_{0.01,9} = 21.67$

$$\frac{9(1.913)^2}{21.67} \le \sigma^2$$

$$1.5199 \le \sigma^2$$

c) 90% lower confidence bound for σ^2

For $\alpha = 0.1$ and n = 10, $\chi^2_{\alpha,n-1} = \chi^2_{0.1,9} = 14.68$

$$\frac{9(1.913)^2}{14.68} \le \sigma^2$$

$$2.2436 \le \sigma^2$$

$$1.498 \le \sigma$$

d) The lower confidence bound of the 99% two-sided interval is less than the one-sided interval. The lower confidence bound for σ^2 is in part (c) is greater because the confidence is lower.

8-57 95% two sided confidence interval for σ, $n = 39 \quad s = 0.6295$

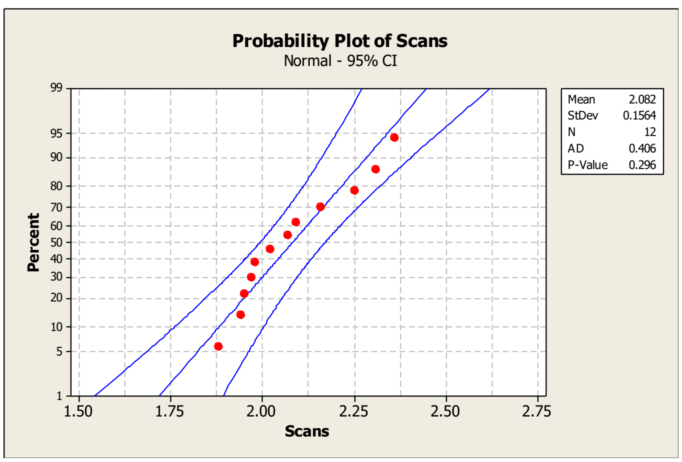
$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,38} = 55.896 \text{ and } \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,38} = 22.878$$

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$
$$\frac{38(0.6295)^2}{55.896} \le \sigma^2 \le \frac{38(0.6295)^2}{22.878}$$
$$0.265 \le \sigma^2 \le 0.658$$
$$0.514 < \sigma < 0.811$$

8-58 a) 95% two sided confidence interval for σ, $n = 12 \quad s = 0.1564$

$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,11} = 21.920 \text{ and } \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,11} = 3.816$$
$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$
$$\frac{11(0.1564)^2}{21.920} \le \sigma^2 \le \frac{11(0.1564)^2}{3.816}$$
$$0.012 \le \sigma^2 \le 0.070$$
$$0.111 < \sigma < 0.265$$

b) A normal probability plot can be used to check the normality assumption.
The following plot does not indicate serious departures from a normal distribution.



Section 8-4

8-59 a) 95% Confidence Interval on the fraction defective produced with this tool.

$$\hat{p} = \frac{13}{300} = 0.04333 \qquad n = 300 \qquad z_{\alpha/2} = 1.96$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.04333 - 1.96\sqrt{\frac{0.04333(0.95667)}{300}} \le p \le 0.04333 + 1.96\sqrt{\frac{0.04333(0.95667)}{300}}$$

$$0.02029 \le p \le 0.06637$$

b) 95% upper confidence bound $\underline{z_\alpha = z_{0.05} = 1.65}$

$$p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$p \le 0.04333 + 1.650\sqrt{\frac{0.04333(0.95667)}{300}}$$

$$p \le 0.06273$$

8-60 a) 95% Confidence Interval on the proportion of such tears that will heal.

$\hat{p} = 0.676 \quad n = 37 \; z_{\alpha/2} = 1.96$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.676 - 1.96\sqrt{\frac{0.676(0.324)}{37}} \le p \le 0.676 + 1.96\sqrt{\frac{0.676(0.324)}{37}}$$

$$0.5245 \le p \le 0.827$$

b) 95% lower confidence bound on the proportion of such tears that will heal.

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

$$0.676 - 1.64\sqrt{\frac{0.676(0.33)}{37}} \le p$$

$$0.549 \le p$$

8-61 a) 95% confidence interval for the proportion of college graduates in Ohio that voted for George Bush.

$\hat{p} = \frac{412}{768} = 0.536 \quad n = 768 \; z_{\alpha/2} = 1.96$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.536 - 1.96\sqrt{\frac{0.536(0.464)}{768}} \le p \le 0.536 + 1.96\sqrt{\frac{0.536(0.464)}{768}}$$

$$0.501 \le p \le 0.571$$

b) 95% lower confidence bound on the proportion of college graduates in Ohio that voted for George Bush.

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

$$0.536 - 1.64\sqrt{\frac{0.536(0.464)}{768}} \le p$$

$$0.506 \le p$$

8-62 a) 95% Confidence Interval on the death rate from lung cancer.

$$\hat{p} = \frac{823}{1000} = 0.823 \quad n = 1000 \; z_{\alpha/2} = 1.96$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.823 - 1.96\sqrt{\frac{0.823(0.177)}{1000}} \le p \le 0.823 + 1.96\sqrt{\frac{0.823(0.177)}{1000}}$$

$$0.7993 \le p \le 0.8467$$

b) E = 0.03, α = 0.05, $z_{\alpha/2} = z_{0.025} = 1.96$ and $\hat{p} = 0.823$ as the initial estimate of p,

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.03}\right)^2 0.823(1-0.823) = 621.79,$$

n $\cong$ 622.

c) E = 0.03, α = 0.05, $z_{\alpha/2} = z_{0.025} = 1.96$ at least 95% confident

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) = \left(\frac{1.96}{0.03}\right)^2 (0.25) = 1067.11,$$

n $\cong$ 1068.

8-63 a) 95% Confidence Interval on the proportion of rats that are under-weight.

$$\hat{p} = \frac{12}{30} = 0.4 \quad n = 30 \; z_{\alpha/2} = 1.96$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.4 - 1.96\sqrt{\frac{0.4(0.6)}{30}} \le p \le 0.4 + 1.96\sqrt{\frac{0.4(0.6)}{30}}$$

$$0.225 \le p \le 0.575$$

b) E = 0.02, α = 0.05, $z_{\alpha/2} = z_{0.025} = 1.96$ and $\hat{p} = 0.4$as the initial estimate of p,

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.02}\right)^2 0.4(1-0.4) = 2304.96,$$

n $\cong$ 2305.

c) E = 0.02, α = 0.05, $z_{\alpha/2} = z_{0.025} = 1.96$ at least 95% confident

$$n=\left(\frac{z_{\alpha/2}}{E}\right)^2(0.25)=\left(\frac{1.96}{0.02}\right)^2(0.25)=2401.$$

8-64 a) 95% Confidence Interval on the true proportion of helmets showing damage

$$\hat{p}=\frac{18}{50}=0.36 \qquad n=50 \qquad z_{\alpha/2}=1.96$$

$$\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\le p\le \hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.36-1.96\sqrt{\frac{0.36(0.64)}{50}}\le p\le 0.36+1.96\sqrt{\frac{0.36(0.64)}{50}}$$

$$0.227\le p\le 0.493$$

b) $n=\left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)=\left(\frac{1.96}{0.02}\right)^2 0.36(1-0.36)=2212.76$

$n\cong 2213$

c) $n=\left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)=\left(\frac{1.96}{0.02}\right)^2 0.5(1-0.5)=2401$

8-65 The worst case would be for $p = 0.5$, thus with E = 0.05 and $\alpha = 0.01$, $z_{\alpha/2} = z_{0.005} = 2.58$ we obtain a sample size of:

$$n=\left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)=\left(\frac{2.58}{0.05}\right)^2 0.5(1-0.5)=665.64,\quad n\cong 666$$

8-66 E = 0.017, $\alpha = 0.01$, $z_{\alpha/2} = z_{0.005} = 2.58$

$$n=\left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)=\left(\frac{2.58}{0.017}\right)^2 0.5(1-0.5)=5758.13,\quad n\cong 5759$$

8-67 a) $\hat{p}=\frac{466}{500}=0.932 \quad n=500 \quad z_{\alpha/2}=1.96$

$$\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\le p\le \hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.932-1.96\sqrt{\frac{0.932(0.068)}{500}}\le p\le 0.932+1.96\sqrt{\frac{0.932(0.068)}{500}}$$

$$0.910\le p\le 0.945$$

b) E = 0.01, $\alpha = 0.05$, $z_{\alpha/2} = z_{0.025} = 1.96$ and $\hat{p} = 0.932$ as the initial estimate of p,

$$n=\left(\frac{z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p})=\left(\frac{1.96}{0.01}\right)^2 0.932(1-0.932)=2439.48,$$

n ≅ 2440

c) E = 0.01, α = 0.05, $z_{\alpha/2} = z_{0.025} = 1.96$
Here we assume the proportion value that generates the greatest variance; namely *p* = 0.5.

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 (0.25) = \left(\frac{1.96}{0.01}\right)^2 (0.25) = 9623.05,$$

n ≅ 9624

8-68 a) $\hat{p} = \frac{180}{200} = 0.9 \quad n = 200 \; z_{\alpha/2} = 1.96$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.9 - 1.96\sqrt{\frac{0.9(0.1)}{200}} \le p \le 0.9 + 1.96\sqrt{\frac{0.9(0.1)}{200}}$$

$$0.858 \le p \le 0.941$$

b) No, the claim of 93% is within the confidence interval for the true proportion of germinated seeds

8-69 $\hat{p} = \frac{13}{300} = 0.0433 \quad n = 300 \; z_{\alpha/2} = 1.96$

The AC confidence interval follows.

$$\left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right] \le p \le \left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right]$$

$$0.025 \le p \le 0.073$$

The traditional confidence interval follows.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.04333 - 1.96\sqrt{\frac{0.04333(0.95667)}{300}} \le p \le 0.04333 + 1.96\sqrt{\frac{0.04333(0.95667)}{300}}$$

$$0.020 \le p \le 0.066$$

The AC confidence interval is similar to the original one.

8-70 The AC confidence interval follows.

$$\hat{p} = \frac{25}{37} = 0.6757 \quad n = 37 \; z_{\alpha/2} = 1.96$$

$$\left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right] \le p \le \left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right]$$

$$0.514 \le p \le 0.804$$

The traditional confidence interval follows.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.676 - 1.96\sqrt{\frac{0.676(0.324)}{37}} \le p \le 0.676 + 1.96\sqrt{\frac{0.676(0.324)}{37}}$$

$$0.5245 \le p \le 0.827$$

The AC confidence interval is shifted to lower values.

8-71 The AC confidence interval follows.

$$\hat{p} = \frac{466}{500} = 0.932 \quad n = 500 \; z_{\alpha/2} = 1.96$$

$$\left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right] \le p \le \left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right]$$

$$0.906 \le p \le 0.951$$

The traditional confidence interval follows.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.932 - 1.96\sqrt{\frac{0.932(0.068)}{500}} \le p \le 0.932 + 1.96\sqrt{\frac{0.932(0.068)}{500}}$$

$$0.910 \le p \le 0.945$$

The AC confidence interval is similar to the original one.

8-72 The AC confidence interval follows.

$$\hat{p} = \frac{180}{200} = 0.9 \quad n = 200 \; z_{\alpha/2} = 1.96$$

$$\left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right] \le p \le \left[\hat{p} + \frac{z^2{}_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2{}_{\alpha/2}}{4n^2}}\right] \Big/ \left[1 + \frac{z^2{}_{\alpha/2}}{n}\right]$$

$$0.851 \le p \le 0.934$$

The traditional confidence interval follows.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.9 - 1.96\sqrt{\frac{0.9(0.1)}{200}} \le p \le 0.9 + 1.96\sqrt{\frac{0.9(0.1)}{200}}$$

$$0.858 \le p \le 0.941$$

The AC confidence interval is slightly narrower than the original one.

Section 8-6

8-73 95% prediction interval on the life of the next tire given $\bar{x} = 60139.7$ $s = 3645.94$ $n = 16$
for $\alpha=0.05$ $t_{\alpha/2,n-1} = t_{0.025,15} = 2.131$

$$\bar{x} - t_{0.025,15} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.025,15} s\sqrt{1+\frac{1}{n}}$$

$$60139.7 - 2.131(3645.94)\sqrt{1+\frac{1}{16}} \le x_{n+1} \le 60139.7 + 2.131(3645.94)\sqrt{1+\frac{1}{16}}$$

$$52131.1 \le x_{n+1} \le 68148.3$$

The prediction interval is considerably wider than the 95% confidence interval ($58,197.3 \le \mu \le 62,082.07$). This is expected because the prediction interval includes the variability in the parameter estimates as well as the variability in a future observation.

8-74 99% prediction interval on the Izod impact data
$n = 20 \quad \bar{x} = 1.25 \quad s = 0.25 \quad t_{0.005,19} = 2.861$

$$\bar{x} - t_{0.005,19} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.005,19} s\sqrt{1+\frac{1}{n}}$$

$$1.25 - 2.861(0.25)\sqrt{1+\frac{1}{20}} \le x_{n+1} \le 1.25 + 2.861(0.25)\sqrt{1+\frac{1}{20}}$$

$$0.517 \le x_{n+1} \le 1.983$$

The lower bound of the 99% prediction interval is considerably lower than the 99% confidence interval ($1.108 \le \mu \le \infty$). This is expected because the prediction interval needs to include the variability in the parameter estimates as well as the variability in a future observation.

8-75 95% prediction Interval on the volume of syrup of the next beverage dispensed
$\bar{x} = 1.10$ $s = 0.015$ $n = 25$ $t_{\alpha/2,n-1} = t_{0.025,24} = 2.064$

$$\bar{x} - t_{0.025,24} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.025,24} s\sqrt{1+\frac{1}{n}}$$

$$1.10 - 2.064(0.015)\sqrt{1+\frac{1}{25}} \le x_{n+1} \le 1.10 - 2.064(0.015)\sqrt{1+\frac{1}{25}}$$

$$1.068 \le x_{n+1} \le 1.13$$

The prediction interval is wider than the confidence interval: $1.094 \le \mu \le 1.106$

8-76 90% prediction interval the value of the natural frequency of the next beam of this type that will be tested.
given $\bar{x} = 231.67$, $s = 1.53$ For $\alpha = 0.10$ and $n = 5$, $t_{\alpha/2,n-1} = t_{0.05,4} = 2.132$

$$\bar{x} - t_{0.05,4} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.05,4} s\sqrt{1+\frac{1}{n}}$$

$$231.67 - 2.132(1.53)\sqrt{1+\frac{1}{5}} \le x_{n+1} \le 231.67 - 2.132(1.53)\sqrt{1+\frac{1}{5}}$$

$$228.1 \le x_{n+1} \le 235.2$$

The 90% prediction interval is wider than the 90% CI.

8-77 95% Prediction Interval on the volume of syrup of the next beverage dispensed
$n = 20 \quad \bar{x} = 485.8 \quad s = 90.34$ $t_{\alpha/2,n-1} = t_{0.025,19} = 2.093$

$$\bar{x} - t_{0.025,19} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.025,19} s\sqrt{1+\frac{1}{n}}$$
$$485.8 - 2.093(90.34)\sqrt{1+\frac{1}{20}} \le x_{n+1} \le 485.8 - 2.093(90.34)\sqrt{1+\frac{1}{20}}$$
$$292.049 \le x_{n+1} \le 679.551$$

The 95% prediction interval is wider than the 95% confidence interval.

8-78 99% prediction interval on the polyunsaturated fat
$n = 6 \quad \bar{x} = 16.98 \quad s = 0.319 \quad t_{0.005,5} = 4.032$

$$\bar{x} - t_{0.005,5} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.005,5} s\sqrt{1+\frac{1}{n}}$$
$$16.98 - 4.032(0.319)\sqrt{1+\frac{1}{6}} \le x_{n+1} \le 16.98 + 4.032(0.319)\sqrt{1+\frac{1}{6}}$$
$$15.59 \le x_{n+1} \le 18.37$$

The prediction interval is much wider than the confidence interval $16.455 \le \mu \le 17.505$.

8-79 Given $\bar{x}$ = 317.2 s = 15.7 n = 10 for α=0.05 $t_{\alpha/2,n-1} = t_{0.005,9} = 3.250$

$$\bar{x} - t_{0.005,9} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.005,9} s\sqrt{1+\frac{1}{n}}$$
$$317.2 - 3.250(15.7)\sqrt{1+\frac{1}{10}} \le x_{n+1} \le 317.2 - 3.250(15.7)\sqrt{1+\frac{1}{10}}$$
$$263.7 \le x_{n+1} \le 370.7$$

The prediction interval is wider.

8-80 95% prediction interval on the next rod diameter tested
$n = 15 \quad \bar{x} = 8.23 \quad s = 0.025 \quad t_{0.025,14} = 2.145$

$$\bar{x} - t_{0.025,14} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.025,14} s\sqrt{1+\frac{1}{n}}$$
$$8.23 - 2.145(0.025)\sqrt{1+\frac{1}{15}} \le x_{n+1} \le 8.23 - 2.145(0.025)\sqrt{1+\frac{1}{15}}$$
$$8.17 \le x_{n+1} \le 8.29$$

95% two-sided confidence interval on mean rod diameter is $8.216 \le \mu \le 8.244$

8-81 90% prediction interval on the next specimen of concrete tested
given $\bar{x}$ = 2260 s = 35.57 n = 12 for α = 0.05 and n = 12, $t_{\alpha/2,n-1} = t_{0.05,11} = 1.796$

$$\bar{x} - t_{0.05,11} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.05,11} s\sqrt{1+\frac{1}{n}}$$

$$2260 - 1.796(35.57)\sqrt{1+\frac{1}{12}} \le x_{n+1} \le 2260 + 1.796(35.57)\sqrt{1+\frac{1}{12}}$$

$$2193.5 \le x_{n+1} \le 2326.5$$

8-82 90% prediction interval on wall thickness on the next bottle tested.
Given $\bar{x} = 4.05$ $s = 0.08$ $n = 25$ for $t_{\alpha/2,n-1} = t_{0.05,24} = 1.711$

$$\bar{x} - t_{0.05,24} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.05,24} s\sqrt{1+\frac{1}{n}}$$

$$4.05 - 1.711(0.08)\sqrt{1+\frac{1}{25}} \le x_{n+1} \le 4.05 - 1.711(0.08)\sqrt{1+\frac{1}{25}}$$

$$3.91 \le x_{n+1} \le 4.19$$

8-83 90% prediction interval for enrichment data given $\bar{x} = 2.9$ $s = 0.099$ $n = 12$ for $\alpha = 0.10$ and $n = 12$, $t_{\alpha/2,n-1} = t_{0.05,11} = 1.796$

$$\bar{x} - t_{0.05,12} s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.05,12} s\sqrt{1+\frac{1}{n}}$$

$$2.9 - 1.796(0.099)\sqrt{1+\frac{1}{12}} \le x_{n+1} \le 2.9 + 1.796(0.099)\sqrt{1+\frac{1}{12}}$$

$$2.71 \le x_{n+1} \le 3.09$$

The 90% confidence interval is

$$\bar{x} - t_{0.05,12} s\sqrt{\frac{1}{n}} \le \mu \le \bar{x} + t_{0.05,12} s\sqrt{\frac{1}{n}}$$

$$2.9 - 1.796(0.099)\sqrt{\frac{1}{12}} \le \mu \le 2.9 - 1.796(0.099)\sqrt{\frac{1}{12}}$$

$$2.85 \le \mu \le 2.95$$

The prediction interval is wider than the CI on the population mean with the same confidence.

The 99% confidence interval is

$$\bar{x} - t_{0.005,12} s\sqrt{\frac{1}{n}} \le \mu \le \bar{x} + t_{0.005,12} s\sqrt{\frac{1}{n}}$$

$$2.9 - 3.106(0.099)\sqrt{\frac{1}{12}} \le \mu \le 2.9 + 3.106(0.099)\sqrt{\frac{1}{12}}$$

$$2.81 \le \mu \le 2.99$$

The prediction interval is even wider than the CI on the population mean with greater confidence.

8-84 To obtain a one sided prediction interval, use $t_{\alpha,n-1}$ instead of $t_{\alpha/2,n-1}$
Because we want a 95% one sided prediction interval,
$t_{\alpha/2,n-1} = t_{0.05,24} = 1.711$ and $\bar{x} = 4.05$ $s = 0.08$ $n = 25$

$$\bar{x} - t_{0.05,24} s\sqrt{1+\frac{1}{n}} \le x_{n+1}$$
$$4.05 - 1.711(0.08)\sqrt{1+\frac{1}{25}} \le x_{n+1}$$
$$3.91 \le x_{n+1}$$

The prediction interval bound is lower than the confidence interval bound of 4.023 mm

8-85 95% tolerance interval on the life of the tires that has a 95% CL
Given $\bar{x}$ = 60139.7 s = 3645.94 n = 16 we find k=2.903

$$\bar{x} - ks,\ \bar{x} + ks$$
$$60139.7 - 2.903(3645.94),\ 60139.7 + 2.903(3645.94)$$
$$(49555.54,\ 70723.86)$$

95% confidence interval (58,197.3 ≤ μ ≤ 62,082.07) is narrower than the 95% tolerance interval.

8-86 99% tolerance interval on the Izod impact strength PVC pipe that has a 90% CL
Given $\bar{x}$=1.25, s=0.25 and n=20 we find k=3.368

$$\bar{x} - ks,\ \bar{x} + ks$$
$$1.25 - 3.368(0.25),\ 1.25 + 3.368(0.25)$$
$$(0.408,\ 2.092)$$

The 99% tolerance interval is much wider than the 99% confidence interval on the population mean (1.090 ≤ μ ≤ 1.410).

8-87 95% tolerance interval on the syrup volume that has 90% confidence level
$\bar{x}$ = 1.10 s = 0.015 n = 25 and k=2.474

$$\bar{x} - ks,\ \bar{x} + ks$$
$$1.10 - 2.474(0.015),\ 1.10 + 2.474(0.015)$$
$$(1.06,\ 1.14)$$

8-88 99% tolerance interval on the polyunsaturated fatty acid in this type of margarine that has a confidence level of 95% $\bar{x}$ = 16.98 s = 0.319 n=6 and k = 5.775

$$\bar{x} - ks,\ \bar{x} + ks$$
$$16.98 - 5.775(0.319),\ 16.98 + 5.775(0.319)$$
$$(15.14,\ 18.82)$$

The 99% tolerance interval is much wider than the 99% confidence interval on the population mean (16.46 ≤ μ ≤ 17.51).

8-89 95% tolerance interval on the rainfall that has a confidence level of 95%

$$n = 20 \quad \bar{x} = 485.8 \quad s = 90.34 \quad k = 2.752$$

$$\bar{x} - ks,\ \bar{x} + ks$$
$$485.8 - 2.752(90.34),\ 485.8 + 2.752(90.34)$$
$$(237.184,\ 734.416)$$

The 95% tolerance interval is much wider than the 95% confidence interval on the population mean ($443.52 \le \mu \le 528.08$).

8-90 95% tolerance interval on the diameter of the rods in exercise 8-27 that has a 90% confidence level
$\bar{x}$ = 8.23 s = 0.0.25 n=15 and k=2.713

$$\bar{x} - ks,\ \bar{x} + ks$$
$$8.23 - 2.713(0.025),\ 8.23 + 2.713(0.025)$$
$$(8.16,\ 8.30)$$

The 95% tolerance interval is wider than the 95% confidence interval on the population mean ($8.216 \le \mu \le 8.244$).

8-91 99% tolerance interval on the brightness of television tubes that has a 95% CL
Given $\bar{x}$ = 317.2 s = 15.7 n = 10 we find k = 4.433

$$\bar{x} - ks,\ \bar{x} + ks$$
$$317.2 - 4.433(15.7),\ 317.2 + 4.433(15.7)$$
$$(247.60,\ 386.80)$$

The 99% tolerance interval is much wider than the 95% confidence interval on the population mean

$$301.06 \le \mu \le 333.34$$

8-92 90% tolerance interval on the comprehensive strength of concrete that has a 90% CL
Given $\bar{x}$ = 2260 s = 35.57 n = 12 we find k=2.404

$$\bar{x} - ks,\ \bar{x} + ks$$
$$2260 - 2.404(35.57),\ 2260 + 2.404(35.57)$$
$$(2174.5,\ 2345.5)$$

The 90% tolerance interval is much wider than the 95% confidence interval on the population mean
$2237.3 \le \mu \le 2282.5$

8-93 99% tolerance interval on rod enrichment data that have a 95% CL
Given $\bar{x}$ = 2.9 s = 0.099 n = 12 we find k=4.150

$$\bar{x} - ks,\ \bar{x} + ks$$
$$2.9 - 4.150(0.099),\ 2.9 + 4.150(0.099)$$
$$(2.49,\ 3.31)$$

The 99% tolerance interval is much wider than the 95% CI on the population mean ($2.84 \le \mu \le 2.96$)

8-94 a) 90% tolerance interval on wall thickness measurements that have a 90% CL
Given $\bar{x}$ = 4.05 s = 0.08 n = 25 we find k=2.077

$$\bar{x} - ks,\ \bar{x} + ks$$
$$4.05 - 2.077(0.08),\ 4.05 + 2.077(0.08)$$
$$(3.88,\ 4.22)$$

The lower bound of the 90% tolerance interval is much lower than the lower bound on the 95% confidence interval on the population mean ($4.023 \le \mu \le \infty$)

b) 90% lower tolerance bound on bottle wall thickness that has confidence level 90%.
given $\bar{x} = 4.05$ $s = 0.08$ $n = 25$ and $k = 1.702$

$$\bar{x} - ks = 4.05 - 1.702(0.08) = 3.91$$

The lower tolerance bound is of interest if we want the wall thickness to be greater than a certain value so that a bottle will not break.

Supplemental Exercises

8-95 Where $\alpha_1 + \alpha_2 = \alpha$. Let $\alpha = 0.05$
Interval for $\alpha_1 = \alpha_2 = \alpha/2 = 0.025$
The confidence level for $\bar{x} - 1.96\sigma/\sqrt{n} \le \mu \le \bar{x} + 1.96\sigma/\sqrt{n}$ is determined by the by the value of z_0 which is 1.96.
From Table III, we find $\Phi(1.96) = P(Z<1.96) = 0.975$ and the confidence level is 95%.
Interval for $\alpha_1 = 0.01,\quad \alpha_2 = 0.04$
The confidence interval is $\bar{x} - 2.33\sigma/\sqrt{n} \le \mu \le \bar{x} + 1.75\sigma/\sqrt{n}$, the confidence level is the same because $\alpha = 0.05$. The symmetric interval does not affect the level of significance; however, it does affect the width. The symmetric interval is narrower.

8-96 $\mu = 50$ σ unknown
a) $n = 16$ $\bar{x} = 52$ $s = 1.5$

$$t_o = \frac{52 - 50}{8/\sqrt{16}} = 1$$

The *P*-value for $t_0 = 1$, degrees of freedom = 15, is between 0.1 and 0.25. Thus, we conclude that the results are not very unusual.

b) n = 30

$$t_o = \frac{52 - 50}{8/\sqrt{30}} = 1.37$$

The P-value for $t_0 = 1.37$, degrees of freedom = 29, is between 0.05 and 0.1. Thus, we conclude that the results are somewhat unusual.

c) n = 100 (with n > 30, the standard normal table can be used for this problem)

$$z_o = \frac{52 - 50}{8/\sqrt{100}} = 2.5$$

The P-value for $z_0 = 2.5$, is 0.00621. Thus we conclude that the results are very unusual.

d) For constant values of $\bar{x}$ and s, increasing only the sample size, we see that the standard error of $\bar{X}$ decreases and consequently a sample mean value of 52 when the true mean is 50 is more unusual for the larger sample sizes.

8-97 $\mu = 50, \quad \sigma^2 = 5$

a) For $n = 16$ find $P(s^2 \geq 7.44)$ or $P(s^2 \leq 2.56)$

$$P(S^2 \geq 7.44) = P\left(\chi^2_{15} \geq \frac{15(7.44)}{5^2}\right) = 0.05 \leq P\left(\chi^2_{15} \geq 22.32\right) \leq 0.10$$

Using computer software $P(S^2 \geq 7.44)$ = 0.0997

$$P(S^2 \leq 2.56) = P\left(\chi^2_{15} \leq \frac{15(2.56)}{5}\right) = 0.05 \leq P\left(\chi^2_{15} \leq 7.68\right) \leq 0.10$$

Using computer software $P(S^2 \leq 2.56)$ = 0.064

b) For $n = 30$ find $P(S^2 \geq 7.44)$ or $P(S^2 \leq 2.56)$

$$P(S^2 \geq 7.44) = P\left(\chi^2_{29} \geq \frac{29(7.44)}{5}\right) = 0.025 \leq P\left(\chi^2_{29} \geq 43.15\right) \leq 0.05$$

Using computer software $P(S^2 \geq 7.44)$ = 0.044

$$P(S^2 \leq 2.56) = P\left(\chi^2_{29} \leq \frac{29(2.56)}{5}\right) = 0.01 \leq P\left(\chi^2_{29} \leq 14.85\right) \leq 0.025$$

Using computer software $P(S^2 \leq 2.56)$ = 0.014.

c) For $n = 71$ $P(s^2 \geq 7.44)$ or $P(s^2 \leq 2.56)$

$$P(S^2 \geq 7.44) = P\left(\chi^2_{70} \geq \frac{70(7.44)}{5}\right) = 0.005 \leq P\left(\chi^2_{70} \geq 104.16\right) \leq 0.01$$

Using computer software $P(S^2 \geq 7.44)$ =0.0051

$$P(S^2 \leq 2.56) = P\left(\chi^2_{70} \leq \frac{70(2.56)}{5}\right) = P\left(\chi^2_{70} \leq 35.84\right) \leq 0.005$$

Using computer software $P(S^2 \leq 2.56)$ < 0.001

d) The probabilities decrease as n increases. As n increases, the sample variance should approach the population variance; therefore, the likelihood of obtaining a sample variance much larger than the population variance decreases.

e) The probabilities decrease as n increases. As n increases, the sample variance should approach the population variance; therefore, the likelihood of obtaining a sample variance much smaller than the population variance decreases.

8-98 a) The data appear to follow a normal distribution based on the normal probability plot because the data fall along a straight line.

b) It is important to check for normality of the distribution underlying the sample data because the confidence intervals to be constructed have the assumption of normality (especially since the sample size is less than 30 and the central limit theorem does not apply).

c) No, with 95% confidence, we cannot infer that the true mean if14.05 because this value is not contained within the given 95% confidence interval.

d) As with part b, to construct a confidence interval on the variance, the normality assumption must hold for the results to be reliable.

e) Yes, it is reasonable to infer that the variance could be 0.35 since the 95% confidence interval on the variance contains this value.

f) i) & ii) No, doctors and children would represent two completely different populations not represented by the population of Canadian Olympic hockey players. Because neither doctors nor children were the target of this study or part of the sample taken, the results should not be extended to these groups.

8-99 a) The probability plot shows that the data appear to be normally distributed.

b) 99% lower confidence bound on the mean $\vec{x} = 25.12, \quad s = 8.42, \quad n = 9 \qquad t_{0.01,8} = 2.896$

$$\bar{x} - t_{0.01,8}\left(\frac{s}{\sqrt{n}}\right) \le \mu$$

$$25.12 - 2.896\left(\frac{8.42}{\sqrt{9}}\right) \le \mu$$

$$16.99 \le \mu$$

The lower bound on the 99% confidence interval shows that the mean comprehensive strength is greater than 16.99 Megapascals with high confidence.

c) 98% two-sided confidence interval on the mean $\vec{x} = 25.12, \quad s = 8.42, \quad n = 9 \quad t_{0.01,8} = 2.896$

$$\bar{x} - t_{0.01,8}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.01,8}\left(\frac{s}{\sqrt{n}}\right)$$

$$25.12 - 2.896\left(\frac{8.42}{\sqrt{9}}\right) \le \mu \le 25.12 + 2.896\left(\frac{8.42}{\sqrt{9}}\right)$$

$$16.99 \le \mu \le 33.25$$

The 98% two-sided confidence interval shows that the mean comprehensive strength is greater than 16.99 Megapascals and less than 33.25 Megapascals with high confidence.

The lower bound of the 99% one sided CI is the same as the lower bound of the 98% two-sided CI because the value of α for the one-sided example is one-half the value for the two-sided example.

d) 99% one-sided upper bound on the confidence interval on σ^2 comprehensive strength

$$s = 8.42, \quad s^2 = 70.90 \quad \chi^2_{0.99,8} = 1.65$$

$$\sigma^2 \le \frac{8(8.42)^2}{1.65}$$

$$\sigma^2 \le 343.74$$

The upper bound on the 99% confidence interval on the variance shows that the variance of the comprehensive strength is less than 343.74 Megapascals^2 with high confidence.

e) 98% two-sided confidence interval on σ^2 of comprehensive strength

$$s = 8.42, \quad s^2 = 70.90 \quad \chi^2_{0.01,9} = 20.09 \quad \chi^2_{0.99,8} = 1.65$$

$$\frac{8(8.42)^2}{20.09} \le \sigma^2 \le \frac{8(8.42)^2}{1.65}$$
$$28.23 \le \sigma^2 \le 343.74$$

The 98% two-sided confidence-interval on the variance shows that the variance of the comprehensive strength is less than 343.74 Megapascals^2 and greater than 28.23 Megapascals^2 with high confidence.

The upper bound of the 99% one-sided CI is the same as the upper bound of the 98% two-sided CI because the value of α for the one-sided example is one-half the value for the two-sided example.

f) 98% two-sided confidence interval on the mean $\vec{x} = 23$, $\quad s = 6.31$, $\quad n = 9 \quad t_{0.01,8} = 2.896$

$$\bar{x} - t_{0.01,8}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.01,8}\left(\frac{s}{\sqrt{n}}\right)$$
$$23 - 2.896\left(\frac{6.31}{\sqrt{9}}\right) \le \mu \le 23 + 2.896\left(\frac{6.31}{\sqrt{9}}\right)$$
$$16.91 \le \mu \le 29.09$$

98% two-sided confidence interval on σ^2 comprehensive strength

$$s = 6.31, \quad s^2 = 39.8 \quad \chi^2_{0.01,9} = 20.09 \quad \chi^2_{0.99,8} = 1.65$$
$$\frac{8(39.8)}{20.09} \le \sigma^2 \le \frac{8(39.8)}{1.65}$$
$$15.85 \le \sigma^2 \le 192.97$$

Fixing the mistake decreased the values of the sample mean and the sample standard deviation. Because the sample standard deviation was decreased, the widths of the confidence intervals were also decreased.

g) A 98% two-sided confidence interval on the mean $\vec{x} = 25$, $\quad s = 8.41$, $\quad n = 9 \quad t_{0.01,8} = 2.896$

$$\bar{x} - t_{0.01,8}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.01,8}\left(\frac{s}{\sqrt{n}}\right)$$
$$25 - 2.896\left(\frac{8.41}{\sqrt{9}}\right) \le \mu \le 25 + 2.896\left(\frac{8.41}{\sqrt{9}}\right)$$
$$16.88 \le \mu \le 33.12$$

98% two-sided confidence interval on σ^2 of comprehensive strength

$$s = 8.41, \quad s^2 = 70.73 \quad \chi^2_{0.01,9} = 20.09 \quad \chi^2_{0.99,8} = 1.65$$
$$\frac{8(8.41)^2}{20.09} \le \sigma^2 \le \frac{8(8.41)^2}{1.65}$$
$$28.16 \le \sigma^2 \le 342.94$$

Fixing the mistake did not affect the sample mean or the sample standard deviation. They are very close to the original values. The widths of the confidence intervals are also very similar.

h) When a mistaken value is near the sample mean, the mistake does not affect the sample mean, standard deviation or confidence intervals greatly. However, when the mistake is not near the sample mean, the value can greatly affect the sample mean, standard deviation and confidence intervals. The farther from the mean, the greater is the effect.

8-100

With σ = 8, the 95% confidence interval on the mean has length of at most 5; the error is then E = 2.5.

a) $n = \left(\frac{z_{0.025}}{2.5}\right)^2 8^2 = \left(\frac{1.96}{2.5}\right)^2 64 = 39.34$ = 40

b) $n = \left(\frac{z_{0.025}}{2.5}\right)^2 6^2 = \left(\frac{1.96}{2.5}\right)^2 36 = 22.13$ = 23

As the standard deviation decreases, with all other values held constant, the sample size necessary to maintain the acceptable level of confidence, and the width of the interval, decreases.

8-101 $\bar{x} = 15.33 \quad s = 0.62 \quad n = 20 \quad k = 2.564$

a) 95% Tolerance Interval of hemoglobin values with 90% confidence

$$\bar{x} - ks,\ \bar{x} + ks$$
$$15.33 - 2.564(0.62),\ 15.33 + 2.564(0.62)$$
$$(13.74,\quad 16.92)$$

b) 99% Tolerance Interval of hemoglobin values with 90% confidence $k = 3.368$

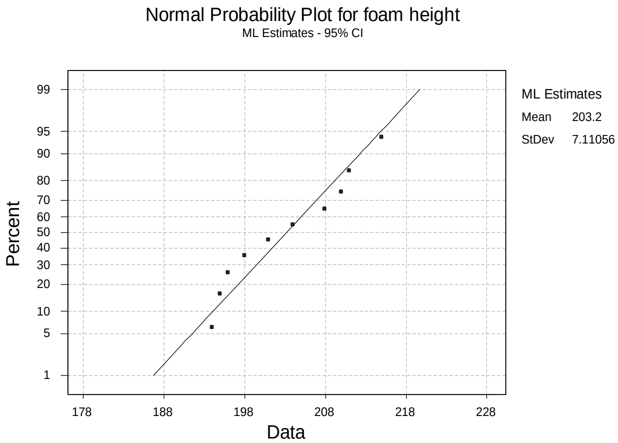
$$\bar{x} - ks,\ \bar{x} + ks$$
$$15.33 - 3.368(0.62),\ 15.33 + 3.368(0.62)$$
$$(13.24, 17.42)$$

8-102 95% prediction interval for the next sample of concrete that will be tested.

Given $\bar{x}$ = 25.12 s = 8.42 n = 9 for α = 0.05 and n = 9, $t_{\alpha/2,n-1} = t_{0.025,8} = 2.306$

$$\bar{x} - t_{0.025,8}s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.025,8}s\sqrt{1+\frac{1}{n}}$$
$$25.12 - 2.306(8.42)\sqrt{1+\frac{1}{9}} \le x_{n+1} \le 25.12 + 2.306(8.42)\sqrt{1+\frac{1}{9}}$$
$$4.65 \le x_{n+1} \le 45.59$$

8-103 a) The data appear to be normally distributed.



b) 95% confidence interval on the mean $\vec{x} = 203.20, \quad s = 7.5, \quad n = 10 \quad t_{0.025,9} = 2.262$

$$\bar{x} - t_{0.025,9}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} - t_{0.025,9}\left(\frac{s}{\sqrt{n}}\right)$$

$$203.2 - 2.262\left(\frac{7.50}{\sqrt{10}}\right) \le \mu \le 203.2 + 2.262\left(\frac{7.50}{\sqrt{10}}\right)$$

$$197.84 \le \mu \le 208.56$$

c) 95% prediction interval on a future sample

$$\bar{x} - t_{0.025,9}s\sqrt{1+\frac{1}{n}} \le \mu \le \bar{x} - t_{0.025,9}s\sqrt{1+\frac{1}{n}}$$

$$203.2 - 2.262(7.50)\sqrt{1+\frac{1}{10}} \le \mu \le 203.2 + 2.262(7.50)\sqrt{1+\frac{1}{10}}$$

$$185.41 \le \mu \le 220.99$$

d) 95% tolerance interval on foam height with 99% confidence $k = 4.265$

$$\bar{x} - ks, \ \bar{x} + ks$$

$$203.2 - 4.265(7.5), \ 203.2 + 4.265(7.5)$$

$$(171.21, \ 235.19)$$

e) The 95% CI on the population mean is the narrowest interval. For the CI, 95% of such intervals contain the population mean. For the prediction interval, 95% of such intervals will cover a future data value. This interval is wider than the CI on the mean. The tolerance interval is the widest interval of all. For the tolerance interval, 99% of such intervals will include 95% of the true distribution of foam height.

8-104

a) Normal probability plot for the coefficient of restitution.

b) 99% CI on the true mean coefficient of restitution

$\bar{x} = 0.624$, $s = 0.013$, $n = 40$ $t_{a/2,\, n-1} = t_{0.005,\, 39} = 2.7079$

$$\bar{x} - t_{0.005,39}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{0.005,39}\frac{s}{\sqrt{n}}$$

$$0.624 - 2.7079\frac{0.013}{\sqrt{40}} \le \mu \le 0.624 + 2.7079\frac{0.013}{\sqrt{40}}$$

$$0.618 \le \mu \le 0.630$$

c) 99% prediction interval on the coefficient of restitution for the next baseball that will be tested.

$$\bar{x} - t_{0.005,39}s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.005,39}s\sqrt{1+\frac{1}{n}}$$

$$0.624 - 2.7079(0.013)\sqrt{1+\frac{1}{40}} \le x_{n+1} \le 0.624 + 2.7079(0.013)\sqrt{1+\frac{1}{40}}$$

$$0.588 \le x_{n+1} \le 0.660$$

d) 99% tolerance interval on the coefficient of restitution with a 95% level of confidence

$$(\bar{x} - ks, \bar{x} + ks)$$

$$(0.624 - 3.213(0.013), 0.624 + 3.213(0.013))$$

$$(0.582, 0.666)$$

e) The confidence interval in part (b) is for the population mean and we may interpret this to imply that 99% of such intervals will cover the true population mean. For the prediction interval, 99% of such intervals will cover a future baseball's coefficient of restitution. For the tolerance interval, 95% of such intervals will cover 99% of the true distribution.

8-105 95% Confidence Interval on the proportion of baseballs with a coefficient of restitution that exceeds 0.635.

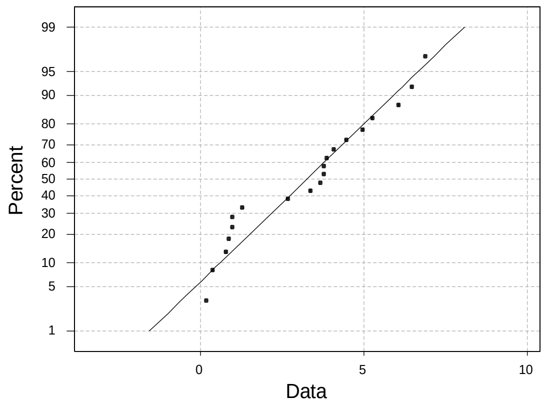
$$\hat{p} = \frac{8}{40} = 0.2 \qquad n = 40 \qquad z_\alpha = 1.65$$

$$\hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

$$0.2 - 1.65\sqrt{\frac{0.2(0.8)}{40}} \le p$$

$$0.0956 \le p$$

8-106 a) The normal probability shows that the data are mostly follow the straight line, however, there are some points that deviate from the line near the middle.



b) 95% CI on the mean dissolved oxygen concentration

$\bar{x}$ = 3.265, s = 2.127, n = 20 $t_{a/2,\, n\text{-}1} = t_{0.025,\, 19}$ = 2.093

$$\bar{x} - t_{0.025,19}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{0.025,19}\frac{s}{\sqrt{n}}$$

$$3.265 - 2.093\frac{2.127}{\sqrt{20}} \le \mu \le 3.265 + 2.093\frac{2.127}{\sqrt{20}}$$

$$2.270 \le \mu \le 4.260$$

c) 95% prediction interval on the oxygen concentration for the next stream in the system that will be tested

$$\bar{x} - t_{0.025,19}s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.025,19}s\sqrt{1+\frac{1}{n}}$$

$$3.265 - 2.093(2.127)\sqrt{1+\frac{1}{20}} \le x_{n+1} \le 3.265 + 2.093(2.127)\sqrt{1+\frac{1}{20}}$$

$$-1.297 \le x_{n+1} \le 7.827$$

d) 95% tolerance interval on the values of the dissolved oxygen concentration with a 99% level of confidence

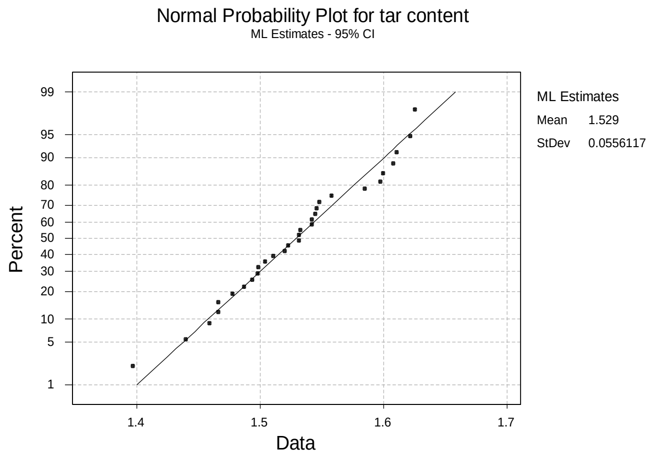
$$(\bar{x} - ks,\ \bar{x} + ks)$$

$$(3.265 - 3.168(2.127),\ 3.265 + 3.168(2.127))$$

$$(-3.473,\ 10.003)$$

e) The confidence interval in part (b) is for the population mean and we may interpret this to imply that 95% of such intervals will cover the true population mean. For the prediction interval, 95% of such intervals will cover a future oxygen concentration. For the tolerance interval, 99% of such intervals will cover 95% of the true distribution

8-107 a) The data appear normally distributed. The data points appear to fall along the normal probability line.



b) 99% CI on the mean tar content

$\bar{x}$ = 1.529, s = 0.0566, n = 30 $t_{a/2, n-1} = t_{0.005, 29}$ = 2.756

$$\bar{x} - t_{0.005,29}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{0.005,29}\frac{s}{\sqrt{n}}$$

$$1.529 - 2.756\frac{0.0566}{\sqrt{30}} \le \mu \le 1.529 + 2.756\frac{0.0566}{\sqrt{30}}$$

$$1.501 \le \mu \le 1.557$$

c) 99% prediction interval on the tar content for the next sample that will be tested..

$$\bar{x} - t_{0.005,19}s\sqrt{1+\frac{1}{n}} \le x_{n+1} \le \bar{x} + t_{0.005,19}s\sqrt{1+\frac{1}{n}}$$

$$1.529 - 2.756(0.0566)\sqrt{1+\frac{1}{30}} \le x_{n+1} \le 1.529 + 2.756(0.0566)\sqrt{1+\frac{1}{30}}$$

$$1.370 \le x_{n+1} \le 1.688$$

d) 99% tolerance interval on the values of the tar content with a 95% level of confidence

$$(\bar{x} - ks, \bar{x} + ks)$$

$$(1.529 - 3.350(0.0566), 1.529 + 3.350(0.0566))$$

$$(1.339, 1.719)$$

e) The confidence interval in part (b) is for the population mean and we may interpret this to imply that 95% of such intervals will cover the true population mean. For the prediction interval, 95% of such intervals will cover a future observed tar content. For the tolerance interval, 99% of such intervals will cover 95% of the true distribution.

8-108 a) 95% Confidence Interval on the population proportion

n=1200 x=8 $\hat{p} = 0.0067$ $z_{\alpha/2}=z_{0.025}=1.96$

$$\hat{p} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.0067 - 1.96\sqrt{\frac{0.0067(1-0.0067)}{1200}} \le p \le 0.0067 + 1.96\sqrt{\frac{0.0067(1-0.0067)}{1200}}$$

$$0.0021 \le p \le 0.0113$$

b) No, there is not sufficient evidence to support the claim that the fraction of defective units produced is one percent or less at $\alpha = 0.05$. This is because the upper limit of the control limit is greater than 0.01.

8-109 99% Confidence Interval on the population proportion

n=1600 x=8 $\hat{p} = 0.005$ $z_{\alpha/2}=z_{0.005}=2.58$

$$\hat{p} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.005 - 2.58\sqrt{\frac{0.005(1-0.005)}{1600}} \le p \le 0.005 + 2.58\sqrt{\frac{0.005(1-0.005)}{1600}}$$

$$0.0004505 \le p \le 0.009549$$

b) E = 0.008, α = 0.01, $z_{\alpha/2} = z_{0.005} = 2.58$

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) = \left(\frac{2.58}{0.008}\right)^2 0.005(1-0.005) = 517.43, \quad n \cong 518$$

c) E = 0.008, α = 0.01, $z_{\alpha/2} = z_{0.005} = 2.58$

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) = \left(\frac{2.58}{0.008}\right)^2 0.5(1-0.5) = 26001.56, \quad n \cong 26002$$

d) A bound on the true population proportion reduces the required sample size by a substantial amount. A sample size of 518 is much smaller than a sample size of over 26,000.

8-110 $\hat{p} = \frac{117}{484} = 0.242$

a) 90% confidence interval; $z_{\alpha/2} = 1.645$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.210 \le p \le 0.274$$

With 90% confidence, the true proportion of new engineering graduates who were planning to continue studying for an advanced degree is between 0.210 and 0.274.

b) 95% confidence interval; $z_{\alpha/2} = 1.96$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.204 \le p \le 0.280$$

With 95% confidence, the true proportion of new engineering graduates who were planning to continue studying for an advanced degree lies between 0.204 and 0.280.

c) Comparison of parts (a) and (b):

The 95% confidence interval is wider than the 90% confidence interval. Higher confidence produces wider intervals, all other values held constant.

d) Yes, since both intervals contain the value 0.25, thus there in not enough evidence to conclude that the true proportion differs from 0.25.

8-111

a) The data appear to follow a normal distribution based on the normal probability plot. The data fall along a straight line.

b) It is important to check for normality of the distribution underlying the sample data because the confidence intervals have the assumption of normality (especially since the sample size is less than 30 and the central limit theorem does not apply).

c) 95% confidence interval for the mean

$n = 11 \quad \bar{x} = 22.73 \quad s = 6.33 \quad t_{0.025,10} = 2.228$

$$\bar{x} - t_{0.025,10}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,10}\left(\frac{s}{\sqrt{n}}\right)$$

$$22.73 - 2.228\left(\frac{6.33}{\sqrt{11}}\right) \le \mu \le 22.73 + 2.228\left(\frac{6.33}{\sqrt{11}}\right)$$

$$18.478 \le \mu \le 26.982$$

d) As with part b, to construct a confidence interval on the variance, the normality assumption must hold for the results to be reliable.

e) 95% confidence interval for variance

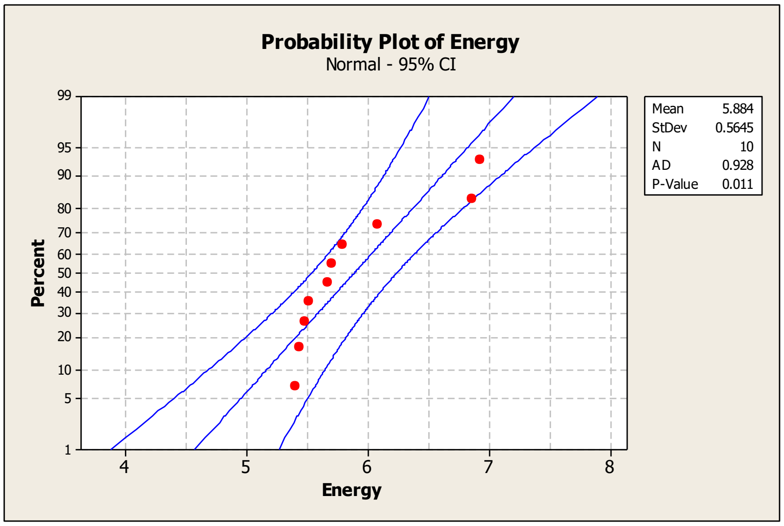
$n = 11 \quad s = 6.33$

$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,10} = 20.48$ and $\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,10} = 3.25$

$$\frac{10(6.33)^2}{20.48} \le \sigma^2 \le \frac{10(6.33)^2}{3.25}$$

$$19.565 \le \sigma^2 \le 123.289$$

8-112 a) The data appear to be normally distributed based on examination of the normal probability plot below.



b) 99% upper confidence interval on mean energy (BMR)

$n = 10 \quad \bar{x} = 5.884 \quad s = 0.5645 \quad t_{0.005,9} = 3.250$

$$\bar{x} - t_{0.005,9}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.005,9}\left(\frac{s}{\sqrt{n}}\right)$$

$$5.884 - 3.250\left(\frac{0.5645}{\sqrt{10}}\right) \le \mu \le 5.884 + 3.250\left(\frac{0.5645}{\sqrt{10}}\right)$$

$$5.304 \le \mu \le 6.464$$

Mind Expanding Exercises

8-113 a) $P(\chi^2_{1-\frac{\alpha}{2},2r} < 2\lambda T_r < \chi^2_{\frac{\alpha}{2},2r}) = 1-\alpha$

$$= P\left(\frac{\chi^2_{1-\frac{\alpha}{2},2r}}{2T_r} < \lambda < \frac{\chi^2_{\frac{\alpha}{2},2r}}{2T_r}\right)$$

Then a confidence interval for $\mu = \dfrac{1}{\lambda}$ is $\left(\dfrac{2T_r}{\chi^2_{\frac{\alpha}{2},2r}}, \dfrac{2T_r}{\chi^2_{1-\frac{\alpha}{2},2r}}\right)$

b) n = 20 , r = 10 , and the observed value of T_r is 199 + 10(29) = 489.

A 95% confidence interval for $\dfrac{1}{\lambda}$ is $\left(\dfrac{2(489)}{34.17}, \dfrac{2(489)}{9.59}\right) = (28.62, 101.98)$

8-114 $\alpha_1 = \displaystyle\int_{z_{\alpha_1}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - \int_{-\infty}^{z_{\alpha_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

Therefore, $1-\alpha_1 = \Phi(z_{\alpha_1})$.

To minimize L we need to minimize $\Phi^{-1}(1-\alpha_1) + \Phi(1-\alpha_2)$ subject to $\alpha_1 + \alpha_2 = \alpha$.

Therefore, we need to minimize $\Phi^{-1}(1-\alpha_1)+\Phi(1-\alpha+\alpha_1)$.

$$\frac{\partial}{\partial\alpha_1}\Phi^{-1}(1-\alpha_1) = -\sqrt{2\pi}\ e^{\frac{z_{\alpha_1}^2}{2}}$$

$$\frac{\partial}{\partial\alpha_1}\Phi^{-1}(1-\alpha+\alpha_1) = \sqrt{2\pi}\ e^{\frac{z_{\alpha-\alpha_1}^2}{2}}$$

Upon setting the sum of the two derivatives equal to zero, we obtain $e^{\frac{z_{\alpha-\alpha_1}^2}{2}} = e^{\frac{z_{\alpha_1}^2}{2}}$. This is solved by $z_{\alpha_1} = z_{\alpha-\alpha_1}$. Consequently, $\alpha_1 = \alpha - \alpha_1$, $2\alpha_1 = \alpha$ and $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.

8-115 a) n = ½ + (1.9/.1)(9.4877/4), then n = 46

b) (10 - 0.5)/(9.4877/4) = (1 + p)/(1 - p)

p = 0.6004 between 10.19 and 10.41.

8-116 a)

$$P(X_i \le \tilde{\mu}) = 1/2$$
$$P(allX_i \le \tilde{\mu}) = (1/2)^n$$
$$P(allX_i \ge \tilde{\mu}) = (1/2)^n$$

$$P(A\cup B) = P(A)+P(B)-P(A\cap B)$$
$$= \left(\frac{1}{2}\right)^n + \left(\frac{1}{2}\right)^n = 2\left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^{n-1}$$
$$1-P(A\cup B) = P(\min(X_i) < \tilde{\mu} < \max(X_i)) = 1-\left(\frac{1}{2}\right)^n$$

b) $P(\min(X_i) < \tilde{\mu} < \max(X_i)) = 1-\alpha$

The confidence interval is min(X_i), max(X_i)

8-117

From the definition of a confidence interval we expect 950 of the confidence intervals to include the value of μ. Let X be the number of intervals that contain the true mean (μ). We can use the large sample approximation to determine the probability that P(930 < X < 970).

Let $p = \frac{950}{1000} = 0.950$ $p_1 = \frac{930}{1000} = 0.930$ and $p_2 = \frac{970}{1000} = 0.970$

The variance is estimated by $\frac{p(1-p)}{n} = \frac{0.950(0.050)}{1000}$

$$P(0.930 < p < 0.970) = P\left(Z < \frac{(0.970 - 0.950)}{\sqrt{\frac{0.950(0.050)}{1000}}}\right) - P\left(Z < \frac{(0.930 - 0.950)}{\sqrt{\frac{0.950(0.050)}{1000}}}\right)$$

$$= P\left(Z < \frac{0.02}{0.006892}\right) - P\left(Z < \frac{-0.02}{0.006892}\right) = P(Z < 2.90) - P(Z < -2.90) = 0.9963$$

8-118 $\tilde{p} = \frac{2}{54} = 0.0370 \quad n = 50 \; z_{\alpha/2} = 1.96$

$$\tilde{p} - z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} \le p \le \tilde{p} + z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$$

$$0.0370 - 1.96\sqrt{\frac{0.0370(1-0.0370)}{50}} \le p \le 0.0370 + 1.96\sqrt{\frac{0.0370(1-0.0370)}{50}}$$

$$-0.015 \le p \le 0.089$$

$$p \le 0.089$$

8-119 a) $\hat{p} = \frac{2}{35} = 0.0571 \quad n = 35 \; z_{\alpha/2} = 1.96$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.0571 - 1.96\sqrt{\frac{0.0571(1-0.0571)}{35}} \le p \le 0.0571 + 1.96\sqrt{\frac{0.0571(1-0.0571)}{35}}$$

$$-0.020 \le p \le 0.134$$

$$p \le 0.134$$

b) $\tilde{p} = \frac{4}{39} = 0.1026 \quad n = 35 \; z_{\alpha/2} = 1.96$

$$\tilde{p} - z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} \le p \le \tilde{p} + z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$$

$$0.1026 - 1.96\sqrt{\frac{0.1026(1-0.1026)}{35}} \le p \le 0.1026 + 1.96\sqrt{\frac{0.1026(1-0.1026)}{35}}$$

$$0.002 \le p \le 0.203$$

This confidence interval is much wider than the interval in part (a). Because of the small sample size and small estimated proportion, the interval in part (a) can be overly optimistic and the modified interval in part (b) is probably a better choice.

CHAPTER 9

Section 9-1

9-1 a) $H_0 : \mu = 25, H_1 : \mu \neq 25$ Yes, because the hypothesis is stated in terms of the parameter of interest, inequality is in the alternative hypothesis, and the value in the null and alternative hypotheses matches.

b) $H_0 : \sigma > 10, H_1 : \sigma = 10$ No, because the inequality is in the null hypothesis.

c) $H_0 : \bar{x} = 50, H_1 : \bar{x} \neq 50$ No, because the hypothesis is stated in terms of the statistic rather than the parameter.

d) $H_0 : p = 0.1, H_1 : p = 0.3$ No, the values in the null and alternative hypotheses do not match and both of the hypotheses are equality statements.

e) $H_0 : s = 30, H_1 : s > 30$ No, because the hypothesis is stated in terms of the statistic rather than the parameter.

9-2

The conclusion does not provide strong evidence that the critical dimension mean equals 100nm. There is not sufficient evidence to reject the null hypothesis.

9-3 a) $H_0 : \sigma = 20nm, H_1 : \sigma < 20nm$

b) This result does not provide strong evidence that the standard deviation has not been reduced. There is insufficient evidence to reject the null hypothesis, but this is not strong support for the null hypothesis.

9-4 a) $H_0 : \mu = 25 Newtons, H_1 : \mu < 25 Newtons$

b) No, this result only implies that we do not have enough evidence to support H_1.

9-5 α = P(reject H_0 when H_0 is true)

= P($\bar{X} \leq 11.5$ when μ = 12) = $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{11.5-12}{0.5/\sqrt{4}}\right)$ = P(Z ≤ −2) = 0.02275.

The probability of rejecting the null hypothesis when it is true is 0.02275.

b) β = P(accept H_0 when μ = 11.25) = $P(\bar{X} > 11.5 | \mu = 11.25)$

= $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{11.5-11.25}{0.5/\sqrt{4}}\right)$ = P(Z > 1.0) = 1 − P(Z ≤ 1.0) = 1 − 0.84134 = 0.15866

The probability of failing to reject the null hypothesis when it is false is 0.15866

c) β = P(accept H_0 when μ = 11.25) =

= $P(\bar{X} > 11.5 | \mu = 11.5)$= $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{11.5-11.5}{0.5/\sqrt{4}}\right)$

= P(Z > 0) = 1 − P(Z ≤ 0) = 1 − 0.5 = 0.5

The probability of failing to reject the null hypothesis when it is false is 0.5

9-6 a) α = P($\bar{X} \leq 11.5 \mid \mu = 12$) = $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{11.5-12}{0.5/\sqrt{16}}\right)$ = P(Z ≤ −4) = 0.

The probability of rejecting the null hypothesis when it is true is approximately 0 with a sample size of 16.

b) β = P($\bar{X} > 11.5 \mid \mu = 11.25$) = $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{11.5-11.25}{0.5/\sqrt{16}}\right)$

= P(Z > 2) = 1 − P(Z ≤ 2)= 1− 0.97725 = 0.02275.

c) The probability of failing to reject the null hypothesis when it is false is 0.02275.

$\beta = P(\overline{X} > 11.5 \mid \mu = 11.5) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > \frac{11.5-11.5}{0.5/\sqrt{16}}\right) = P(Z > 0) = 1 - P(Z \le 0) = 1 - 0.5 = 0.5$

The probability of failing to rejct the null hypothesis when it is false is 0.5.

9-7 The critical value for the one-sided test is

$$\overline{X} \le 12 - z_\alpha 0.5/\sqrt{n}$$

a) $\alpha = 0.01$, $n = 4$, from Table III $-2.33 = z_\alpha$ and $\overline{X} \le 11.42$

b) $\alpha = 0.05$, $n = 4$, from Table III $-1.65 = z_\alpha$ and $\overline{X} \le 11.59$

c) $\alpha = 0.01$, $n = 16$, from Table III $-2.33 = z_\alpha$ and $\overline{X} \le 11.71$

d) $\alpha = 0.05$, $n = 16$, from Table III $-1.65 = z_\alpha$ and $\overline{X} \le 11.95$

9-8 a) $\beta = P(\overline{X} > 11.59 \mid \mu = 11.5) = P(Z > 0.36) = 1 - 0.6406 = 0.3594$

b) $\beta = P(\overline{X} > 11.79 \mid \mu = 11.5) = P(Z > 2.32) = 1 - 0.9898 = 0.0102$

c) Notice that the value of β decreases as n increases

9-9 a) $\bar{x} = 11.25$, then P-value= $P\left(Z \le \frac{11.25-12}{0.5/\sqrt{4}}\right) = P(Z \le -3) = 0.00135$

b) $\bar{x} = 11.0$, then P-value= $P\left(Z \le \frac{11.0-12}{0.5/\sqrt{4}}\right) = P(Z \le -4) \le 0.000033$

c) $\bar{x} = 11.75$, then P-value= $P\left(Z \le \frac{11.75-12}{0.5/\sqrt{4}}\right) = P(Z \le -1) = 0.158655$

9-10 a) $\alpha = P(\overline{X} \le 98.5) + P(\overline{X} > 101.5)$

$$= P\left(\frac{\overline{X}-100}{2/\sqrt{9}} \le \frac{98.5-100}{2/\sqrt{9}}\right) + P\left(\frac{\overline{X}-100}{2/\sqrt{9}} > \frac{101.5-100}{2/\sqrt{9}}\right)$$

$= P(Z \le -2.25) + P(Z > 2.25) = (P(Z \le -2.25)) + (1 - P(Z \le 2.25))$

$= 0.01222 + 1 - 0.98778 = 0.01222 + 0.01222 = 0.02444$

b) $\beta = P(98.5 \le \overline{X} \le 101.5$ when $\mu = 103)$

$$= P\left(\frac{98.5-103}{2/\sqrt{9}} \le \frac{\overline{X}-103}{2/\sqrt{9}} \le \frac{101.5-103}{2/\sqrt{9}}\right)$$

$= P(-6.75 \le Z \le -2.25) = P(Z \le -2.25) - P(Z \le -6.75) = 0.01222 - 0 = 0.01222$

c) $\beta = P(98.5 \le \overline{X} \le 101.5 \mid \mu = 105)$

$$= P\left(\frac{98.5-105}{2/\sqrt{9}} \le \frac{\overline{X}-105}{2/\sqrt{9}} \le \frac{101.5-105}{2/\sqrt{9}}\right)$$

$= P(-9.75 \le Z \le -5.25) = P(Z \le -5.25) - P(Z \le -9.75) = 0 - 0 = 0$

The probability of failing to reject the null hypothesis when it is actually false is smaller in part (c) because the true mean, μ = 105, is further from the acceptance region. That is, there is a greater difference between the true mean and the hypothesized mean.

9-11 Use n = 5, everything else held constant:

a) P($\overline{X} \le 98.5$) + P($\overline{X} > 101.5$)

$$= P\left(\frac{\overline{X}-100}{2/\sqrt{5}} \le \frac{98.5-100}{2/\sqrt{5}}\right) + P\left(\frac{\overline{X}-100}{2/\sqrt{5}} > \frac{101.5-100}{2/\sqrt{5}}\right)$$

= P(Z ≤ –1.68) + P(Z > 1.68)

= P(Z ≤ –1.68) + (1 – P(Z ≤ 1.68))= 0.04648 + (1 – 0.95352) = 0.09296

b) β = P(98.5 ≤ $\overline{X}$ ≤ 101.5 when μ = 103)

$$= P\left(\frac{98.5-103}{2/\sqrt{5}} \le \frac{\overline{X}-103}{2/\sqrt{5}} \le \frac{101.5-103}{2/\sqrt{5}}\right)$$

= P(–5.03 ≤ Z ≤ –1.68) = P(Z ≤ –1.68) – P(Z ≤ –5.03) = 0.04648 – 0 = 0.04648

c) β = P(98.5 ≤ $\overline{x}$ ≤ 101.5 when μ = 105)

$$= P\left(\frac{98.5-105}{2/\sqrt{5}} \le \frac{\overline{X}-105}{2/\sqrt{5}} \le \frac{101.5-105}{2/\sqrt{5}}\right)$$

= P(–7.27≤ Z ≤ –3.91)

= P(Z ≤ –3.91) – P(Z ≤ –7.27) = 0.00005 – 0 = 0.00005

It is smaller because it is not likely to accept the product when the true mean is as high as 105.

9-12 $\mu_0 - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \le \overline{X} \le \mu_0 + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$, where σ =2

a) α = 0.01, n = 9, then $z_{\alpha/2}$ =2,57, then 98.29,101.71

b) α = 0.05, n = 9, then $z_{\alpha/2}$ =1.96, then 98.69,101.31

c) α = 0.01, n = 5, then $z_{\alpha/2}$ = 2.57, then 97.70,102.30

d) α = 0.05, n = 5, then $z_{\alpha/2}$ =1.96, then 98.25,101.75

9-13 δ = 103 - 100 = 3

δ > 0 then $\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$, where σ = 2

a) β = P(98.69< $\overline{X}$ <101.31|μ=103) = P(-6.47<Z<-2.54) = 0.0055

b) β = P(98.25< $\overline{X}$ <101.75|μ=103) = P(-5.31<Z<-1.40) = 0.0808

c) As n increases, β decreases

9-14 a) P-value = 2(1 - $\Phi(|Z_0|)$) = 2(1 - $\Phi\left(\left|\frac{98-100}{2/\sqrt{9}}\right|\right)$) = 2(1 - $\Phi(3)$) = 2(1 - 0.99865) = 0.0027

b) P-value = 2(1 - $\Phi(|Z_0|)$) = 2(1 - $\Phi\left(\left|\frac{101-100}{2/\sqrt{9}}\right|\right)$) = 2(1 - $\Phi(1.5)$) = 2(1 - 0.93319) = 0.13362

c) P-value = 2(1 - $\Phi(|Z_0|)$) = 2(1 - $\Phi\left(\left|\dfrac{102-100}{2/\sqrt{9}}\right|\right)$) = 2(1 - $\Phi(3)$) = 2(1 - 0.99865) = 0.0027

9-15 a) α = P($\overline{X}$ > 185 when μ = 175)

$= P\left(\dfrac{\overline{X}-175}{20/\sqrt{10}} > \dfrac{185-175}{20/\sqrt{10}}\right)$

= P(Z > 1.58) = 1 – P(Z ≤ 1.58) = 1 – 0.94295 = 0.05705

b) β = P($\overline{X}$ ≤ 185 when μ = 185)

$= P\left(\dfrac{\overline{X}-185}{20/\sqrt{10}} \le \dfrac{185-185}{20/\sqrt{10}}\right)$ = P(Z ≤ 0) = 0.5

c) β = P($\overline{X}$ ≤ 185 when μ = 195)

$= P\left(\dfrac{\overline{X}-195}{20/\sqrt{10}} \le \dfrac{185-195}{20/\sqrt{10}}\right)$ = P(Z ≤ –1.58) = 0.05705

9-16 Using n = 16:

a) α = P($\overline{X}$ > 185 when μ = 175)

$= P\left(\dfrac{\overline{X}-175}{20/\sqrt{16}} > \dfrac{185-175}{20/\sqrt{16}}\right)$

= P(Z > 2)= 1 – P(Z ≤ 2) = 1 – 0.97725 = 0.02275

b) β = P($\overline{X}$ ≤ 185 when μ = 185)

$= P\left(\dfrac{\overline{X}-195}{20/\sqrt{16}} \le \dfrac{185-185}{20/\sqrt{16}}\right)$ = P(Z ≤ 0) = 0.5

c) β = P($\overline{X}$ ≤ 185 when μ = 195)

$= P\left(\dfrac{\overline{X}-195}{20/\sqrt{16}} \le \dfrac{185-195}{20/\sqrt{16}}\right)$ = P(Z ≤ –2) = 0.02275

9-17 $\overline{X} \ge 175 + Z_\alpha\left(\dfrac{20}{\sqrt{n}}\right)$

a) α = 0.01, n = 10, then z_α =2.32 and critical value is 189.67

b) α = 0.05, n = 10, then z_α =1.64 and critical value is 185.93

c) α = 0.01, n = 16, then z_α = 2.32 and critical value is 186.6

d) α = 0.05, n = 16, then z_α =1.64 and critical value is 183.2

9-18 a) α = 0.05, n =10, then the critical value 185.93 (from the previous exercise part (b)) β = P($\overline{X}$ ≤ 185.37 when μ = 185)

= $P\left(\frac{\overline{X}-185}{20/\sqrt{10}} \le \frac{185.93-185}{20/\sqrt{10}}\right)$ = P(Z ≤ 0.147) = 0.5584

b) α = 0.05, n =16, then the critical value 183.2 (from the previous exercise part (d)), then

β = P($\overline{X}$ ≤ 183.2 when μ = 185)

= $P\left(\frac{\overline{X}-185}{20/\sqrt{16}} \le \frac{183.2-185}{20/\sqrt{16}}\right)$ = P(Z ≤ -0.36) = 0.3594

c) As n increases, β decreases

9-19 P-value = 1 – $\Phi(Z_0)$) where $Z_0 = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$

a) $\overline{X}$ = 180 then $Z_0 = \frac{180-175}{20/\sqrt{10}} = 0.79$

P-value = 1- $\Phi(0.79) = 1-0.7852 = 0.2148$

b) $\overline{X}$ = 190 then $Z_0 = \frac{190-175}{20/\sqrt{10}} = 2.37$

P-value = 1- $\Phi(2.37) = 1-0.991106 = 0.008894$

c) $\overline{X}$ =170 then $Z_0 = \frac{170-175}{20/\sqrt{10}} = -0.79$

P-value = 1- $\Phi(-0.79) = 1-0.214764 = 0.785236$

9-20 a) α = P($\overline{X}$ ≤ 4.85 when μ = 5) + P($\overline{X}$ > 5.15 when

= $P\left(\frac{\overline{X}-5}{0.25/\sqrt{8}} \le \frac{4.85-5}{0.25/\sqrt{8}}\right) + P\left(\frac{\overline{X}-5}{0.25/\sqrt{8}} > \frac{5.15-5}{0.25/\sqrt{8}}\right)$

= P(Z ≤ –1.7) + P(Z > 1.7) = P(Z ≤ –1.7) + (1 – P(Z ≤ 1.7) = 0.04457 + (1 – 0.95543) = 0.08914

b) Power = 1 – β

β = P(4.85 ≤ $\overline{X}$ ≤ 5.15 when μ = 5.1)

= $P\left(\frac{4.85-5.1}{0.25/\sqrt{8}} \le \frac{\overline{X}-5.1}{0.25/\sqrt{8}} \le \frac{5.15-5.1}{0.25/\sqrt{8}}\right)$

= P(–2.83 ≤ Z ≤ 0.566) = P(Z ≤ 0.566) – P(Z ≤ –2.83) = 0.71566 – 0.00233 = 0.71333

1 – β = 0.2867

9-21 Using n = 16:

a) α = P($\overline{X}$ ≤ 4.85 | μ = 5) + P($\overline{X}$ > 5.15 | μ = 5)

= $P\left(\frac{\overline{X}-5}{0.25/\sqrt{16}} \le \frac{4.85-5}{0.25/\sqrt{16}}\right) + P\left(\frac{\overline{X}-5}{0.25/\sqrt{16}} > \frac{5.15-5}{0.25/\sqrt{16}}\right)$

= P(Z ≤ –2.4) + P(Z > 2.4) = P(Z ≤ –2.4) +(1 – P(Z ≤ 2.4)) = 2(1 – P(Z ≤ 2.4))

= 2(1 – 0.99180) = 2(0.0082) = 0.0164

b) β = P(4.85 ≤ $\overline{X}$ ≤ 5.15 | μ = 5.1)

$= P\left(\frac{4.85-5.1}{0.25/\sqrt{16}} \le \frac{\overline{X}-5.1}{0.25/\sqrt{16}} \le \frac{5.15-5.1}{0.25/\sqrt{16}}\right)$

= P(–4 ≤ Z ≤ 0.8) = P(Z ≤ 0.8) – P(Z ≤ –4)

= 0.78814 – 0 = 0.78814

1 – β = 0.21186

c) With larger sample size, the value of α decreased from approximately 0.089 to 0.016. The power declined modestly from 0.287 to 0.211, while the value for α declined substantially. If the test with n = 16 were conducted at the α value of 0.089, then it would have greater power than the test with n = 8.

9-22 $\sigma = 0.25, \mu_0 = 5$

a) α = 0.01, n = 8 then

a = $\mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$ =5+2.57*.25/ $\sqrt{8}$ =5.22 and

b = $\mu_0 - z_{\alpha/2}\sigma/\sqrt{n}$ =5-2.57*.25/ $\sqrt{8}$ =4.77

b) α = 0.05, n = 8 then

a = $\mu_0 + Z_{\alpha/2} * \sigma/\sqrt{n}$ =5+1.96*.25/ $\sqrt{8}$ =5.1732 and

b = $\mu_0 - z_{\alpha/2}\sigma/\sqrt{n}$ =5-1.96*.25/ $\sqrt{8}$ =4.8267

c) α = 0.01, n =16 then

a= $\mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$ =5+2.57*.25/ $\sqrt{16}$ =5.1606 and

b= $\mu_0 - z_{\alpha/2}\sigma/\sqrt{n}$ =5-2.57*.25/ $\sqrt{16}$ =4.8393

d) α = 0.05, n =16 then

a= $\mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$ =5+1.96*.25/ $\sqrt{16}$ =5.1225 and

b= $\mu_0 - z_{\alpha/2}\sigma/\sqrt{n}$ =5-1.96*.25/ $\sqrt{16}$ =4.8775

9-23 P-value=2(1 - $\Phi(|Z_0|)$) where $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

a) $\bar{x}$ = 5.2 then $z_0 = \frac{5.2-5}{.25/\sqrt{8}} = 2.26$

P-value = 2(1- $\Phi(2.26)) = 2(1-0.988089) = 0.0238$

b) $\bar{x}$ = 4.7 then $z_0 = \frac{4.7-5}{.25/\sqrt{8}} = -3.39$

P-value = 2(1- $\Phi(3.39)) = 2(1-0.99965) = 0.0007$

c) $\bar{x}$ = 5.1 then $z_0 = \frac{5.1-5}{.25/\sqrt{8}} = 1.1313$

P-value = 2(1- $\Phi(1.1313)) = 2(1-0.870762) = 0.2585$

9-24 a) β = P(4.845< $\overline{X}$ <5.155|μ = 5.05) = P(-2.59<Z<1.33) = 0.9034

b) β = P(4.8775< $\overline{X}$ <5.1225|μ = 5.05) = P(-2.76<Z<1.16) = 0.8741

c) As n increases, β decreases

9-25 X ~ bin(15, 0.4) H_0: $p = 0.4$ and H_1: $p \neq 0.4$

$p_1 = 4/15 = 0.267$ $p_2 = 8/15 = 0.533$

Accept Region: $0.267 \le \hat{p} \le 0.533$

Reject Region: $\hat{p} < 0.267$ or $\hat{p} > 0.533$

Use the normal approximation for parts a) and b)

a) When p = 0.4, $\alpha = P(\hat{p} < 0.267) + P(\hat{p} > 0.533)$

$$= P\left(Z < \frac{0.267 - 0.4}{\sqrt{\frac{0.4(0.6)}{15}}} \right) + P\left(Z > \frac{0.533 - 0.4}{\sqrt{\frac{0.4(0.6)}{15}}} \right)$$
$$= P(Z < -1.05) + P(Z > 1.05)$$
$$= P(Z < -1.05) + (1 - P(Z < 1.05))$$
$$= 0.14686 + 0.14686 = 0.29372$$

b) When p = 0.2

$$\beta = P(0.267 \le \hat{p} \le 0.533) = P\left(\frac{0.267 - 0.2}{\sqrt{\frac{0.2(0.8)}{15}}} \le Z \le \frac{0.533 - 0.2}{\sqrt{\frac{0.2(0.8)}{15}}} \right)$$
$$= P(0.65 \le Z \le 3.22)$$
$$= P(Z \le 3.22) - P(Z \le 0.65)$$
$$= 0.99936 - 0.74215 = 0.2572$$

9-26 X ~ Bin(10, 0.3) Implicitly, H_0: $p = 0.3$ and H_1: $p < 0.3$
$n = 10$
Accept region: $\hat{p} > 0.1$

Reject region: $\hat{p} \le 0.1$

Use the normal approximation for parts a), b) and c):

a) When p =0.3 α = $P(\hat{p} < 0.1) = P\left(Z \le \frac{0.1 - 0.3}{\sqrt{\frac{0.3(0.7)}{10}}} \right)$

$$= P(Z \le -1.38) = 0.08379$$

b) When p = 0.2 β = $P(\hat{p} > 0.1) = P\left(Z > \frac{0.1 - 0.2}{\sqrt{\frac{0.2(0.8)}{10}}} \right)$

$$= P(Z > -0.79) = 1 - P(Z < -0.79) = 0.78524$$

c) Power = 1 – β = 1 – 078524 = 0.21476

9-27 The problem statement implies H_0: $p = 0.6$, H_1: $p > 0.6$ and defines an acceptance region as $\hat{p} \le \frac{400}{500} = 0.80$ and rejection region as $\hat{p} > 0.80$

a) $\alpha = P(\hat{p} > 0.80 \mid p = 0.60) = P\left(Z > \frac{0.80 - 0.60}{\sqrt{\frac{0.6(0.4)}{500}}}\right) = P(Z > 9.13) = 1 - P(Z \le 9.13) \approx 0$

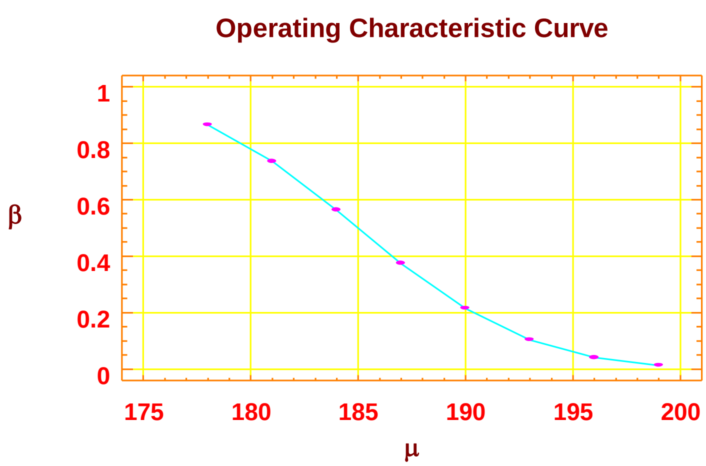
b) $\beta = P(\hat{p} \le 0.8$ when $p = 0.75) = P(Z \le 2.58) = 0.99506$

9-28 a) Operating characteristic curve:

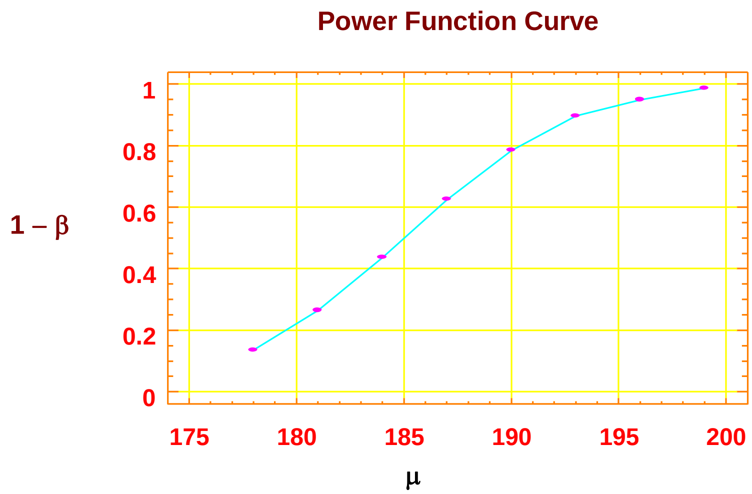
$\bar{x} = 185$

$$\beta = P\left(Z \le \frac{\bar{x} - \mu}{20/\sqrt{10}}\right) = P\left(Z \le \frac{185 - \mu}{20/\sqrt{10}}\right)$$

μ	$P\left(Z \le \frac{185-\mu}{20/\sqrt{10}}\right) =$	β	$1 - \beta$
178	$P(Z \le 1.11) =$	0.8665	0.1335
181	$P(Z \le 0.63) =$	0.7357	0.2643
184	$P(Z \le 0.16) =$	0.5636	0.4364
187	$P(Z \le -0.32) =$	0.3745	0.6255
190	$P(Z \le -0.79) =$	0.2148	0.7852
193	$P(Z \le -1.26) =$	0.1038	0.8962
196	$P(Z \le -1.74) =$	0.0409	0.9591
199	$P(Z \le -2.21) =$	0.0136	0.9864



b)



9-29 a) A type I error means classifying a board as "bad" when it is actually "good".
b) P(Type I error) = 1 − 0.98 = 0.02
c) A type II error means classifying a board as "good" when it is actually "bad".
d) P(Type II error) = 3% = 0.03

9-30

a) Type I error can be improved if a board is only required to pass at least 4 of the 5 tests. With this change, the probability of a type I error (a "good" board classified as "bad") is 0.01.

b) A type II error occurs when a "bad" board is classified as "good". When we decrease the type I error, we improve the rate of detecting "good" boards as "good". However, the probability of a "bad" board passing at least 4 of the 5 tests is 0.03 + 0.20 = 0.23. Therefore, the probability of a type II error increases substantially.

c) The reduction in type I error might be justified when type II error only increases slightly. However, in the example the reduction is not justified. There is a substantial increase in the probability of type II error from 0.3 to 0.23, if the criterion for classifying a good board is changed.

<u>Section 9-2</u>

9-31 a) $H_0: \mu = 10, H_1: \mu > 10$
b) $H_0: \mu = 7, H_1: \mu \neq 7$
c) $H_0: \mu = 5, H_1: \mu < 5$

9-32 a) α = 0.01, then a = $z_{\alpha/2}$ = 2.57 and b = - $z_{\alpha/2}$ = -2.57
b) α = 0.05, then a = $z_{\alpha/2}$ = 1.96 and b = - $z_{\alpha/2}$ = -1.96
c) α = 0.1, then a = $z_{\alpha/2}$ = 1.65 and b = - $z_{\alpha/2}$ = -1.65

9-33 a) α = 0.01, then a = $z_\alpha \cong$ 2.33
b) α = 0.05, then a = $z_\alpha \cong$ 1.64
c) α = 0.1, then a = $z_\alpha \cong$ 1.29

9-34 a) α = 0.01, then a = $z_{1-\alpha} \cong$ -2.33
b) α = 0.05, then a = $z_{1-\alpha} \cong$ -1.64

c) α = 0.1, then a = $z_{1-\alpha}$ ≅ -1.29

9-35 a) P-value = 2(1- $\Phi(|Z_0|)$) = 2(1- $\Phi(2.05)$) ≅ 0.04

b) P-value = 2(1- $\Phi(|Z_0|)$) = 2(1- $\Phi(1.84)$) ≅ 0.066

c) P-value = 2(1- $\Phi(|Z_0|)$) = 2(1- $\Phi(0.4)$) ≅ 0.69

9-36 a) P-value = 1- $\Phi(Z_0)$ = 1- $\Phi(2.05)$ ≅ 0.02

b) P-value = 1- $\Phi(Z_0)$ = 1- $\Phi(-1.84)$ ≅ 0.97

c) P-value = 1- $\Phi(Z_0)$ = 1- $\Phi(0.4)$ ≅ 0.34

9-37 a) P-value = $\Phi(Z_0)$ = $\Phi(2.05)$ ≅ 0.98

b) P-value = $\Phi(Z_0)$ = $\Phi(-1.84)$ ≅ 0.03

c) P-value = $\Phi(Z_0)$ = $\Phi(0.4)$ ≅ 0.65

9-38 a) SE Mean from the sample standard deviation $= \dfrac{\sigma}{\sqrt{N}} = \dfrac{1.475}{\sqrt{25}} = 0.295$

$$z_0 = \frac{35.710 - 35}{1.8/\sqrt{25}} = 1.9722$$

P-value = $2[1-\Phi(Z_0)] = 2[1-\Phi(1.9722)] = 2[1-0.9757] = 0.0486$

Because the P-value < α = 0.05, reject the null hypothesis that μ = 35 at the 0.05 level of significance.

b) A two-sided test because the alternative hypothesis is mu not = 35.

c) 95% CI of the mean is $\bar{x} - z_{0.025}\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{0.025}\dfrac{\sigma}{\sqrt{n}}$

$$35.710 - (1.96)\frac{1.8}{\sqrt{25}} < \mu < 35.710 + (1.96)\frac{1.8}{\sqrt{25}}$$

$$35.0044 < \mu < 36.4156$$

d) P-value = $1-\Phi(Z_0) = 1-\Phi(1.9722) = 1-0.9757 = 0.0243$

9-39 a) StDev = $\sqrt{N}$ (SE Mean) = 0.7495

$$z_0 = \frac{19.889 - 20}{0.75/\sqrt{10}} = -0.468$$

P-value = $1-\Phi(Z_0) = 1-\Phi(-0.468) = 1-0.3199 = 0.6801$

Because the P-value > α = 0.05, we fail to reject the null hypothesis that μ = 20 at the 0.05 level of significance.

b) A one-sided test because the alternative hypothesis is μ > 20

c) 95% CI of the mean is $\bar{x} - z_{0.025}\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{0.025}\dfrac{\sigma}{\sqrt{n}}$

$$19.889 - (1.96)\frac{0.75}{\sqrt{10}} < \mu < 19.889 + (1.96)\frac{0.75}{\sqrt{10}}$$

$$19.4242 < \mu < 20.3539$$

d) P-value = $2[1-\Phi(Z_0)] = 2[1-\Phi(0.468)] = 2[1-0.6801] = 0.6398$

9-40 a) SE Mean from the sample standard deviation $= \dfrac{s}{\sqrt{N}} = \dfrac{1.015}{\sqrt{16}} = 0.2538$

$$z_0 = \frac{15.016 - 14.5}{1.1/\sqrt{16}} = 1.8764$$

P-value= $1 - \Phi(Z_0) = 1 - \Phi(1.8764) = 1 - 0.9697 = 0.0303$

Because the P-value < α = 0.05, reject the null hypothesis that μ = 14.5 at the 0.05 level of significance.

b) A one-sided test because the alternative hypothesis is $\mu > 14.5$

c) 95% lower CI of the mean is $\bar{x} - z_{0.05}\dfrac{\sigma}{\sqrt{n}} \le \mu$

$$15.016 - (1.645)\frac{1.1}{\sqrt{16}} \le \mu$$

$$14.5636 \le \mu$$

d) P-value = $2[1 - \Phi(Z_0)] = 2[1 - \Phi(1.8764)] = 2[1 - 0.9697] = 0.0606$

9-41 a) SE Mean from the sample standard deviation $= \dfrac{s}{\sqrt{N}} = \dfrac{2.365}{\sqrt{12}} = 0.6827$

b) A one-sided test because the alternative hypothesis is $\mu > 99$.

c) If the null hypothesis is changed to the μ = 98, $z_0 = \dfrac{100.039 - 98}{2.5/\sqrt{12}} = 2.8253$

Because $\Phi(2.8253)$ is close to 1, the P-value $= 1 - \Phi(2.8253)$ = 0.002 is very small and close to 0. Thus, the P-value < α = 0.05, and we reject the null hypothesis at the 0.05 level of significance.

d) 95% lower CI of the mean is $\bar{x} - z_{0.05}\dfrac{\sigma}{\sqrt{n}} \le \mu$

$$100.039 - (1.645)\frac{2.5}{\sqrt{12}} \le \mu$$

$$98.8518 \le \mu$$

e) If the alternative hypothesis is changed to the mu $\neq$ 99, $z_0 = \dfrac{100.039 - 99}{2.5/\sqrt{12}} = 1.4397$

P-value = $2[1 - \Phi(Z_0)] = 2[1 - \Phi(1.4397)] = 2[1 - 0.9250] = 0.15$

Because the P-value > α = 0.05, we fail to reject the null hypothesis at the 0.05 level of significance.

9-42 a)

1) The parameter of interest is the true mean water temperature, μ.

2) $H_0 : \mu = 100$

3) $H_1 : \mu > 100$

4) $z_0 = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

5) Reject H_0 if $z_0 > z_\alpha$ where $\alpha = 0.05$ and $z_{0.05} = 1.65$

6) $\bar{x} = 98$, $\sigma = 2$

$$z_0 = \frac{98 - 100}{2/\sqrt{9}} = -3.0$$

7) Because -3.0 < 1.65, fail to reject H_0. The mean water temperature is not significantly greater than 100 at α = 0.05.

b) P-value = $1-\Phi(-3.0)=1-0.00135=0.99865$

c) $\beta=\Phi\left(z_{0.05}+\dfrac{100-104}{2/\sqrt{9}}\right)$ = Φ(1.65 + –6) = Φ(-4.35) ≅0

9-43 a)

1) The parameter of interest is the true mean crankshaft wear, μ.
2) $H_0 : \mu = 3$
3) $H_1 : \mu \neq 3$
4) $z_0 = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
5) Reject H_0 if $z_0 < -z_{\alpha/2}$ where $\alpha = 0.05$ and $-z_{0.025} = -1.96$ or $z_0 > z_{\alpha/2}$ where $\alpha = 0.05$ and $z_{0.025} = 1.96$
6) $\bar{x} = 2.78$, $\sigma = 0.9$

$$z_0=\frac{2.78-3}{0.9/\sqrt{15}}=-0.95$$

7) Because –0.95 > -1.96, fail to reject the null hypothesis. There is not sufficient evidence to support the claim the mean crankshaft wear differs from 3 at α = 0.05.

b) $\beta=\Phi\left(z_{0.025}+\dfrac{3-3.25}{0.9/\sqrt{15}}\right)-\Phi\left(-z_{0.025}+\dfrac{3-3.25}{0.9/\sqrt{15}}\right)$

= Φ(1.96 + –1.08) – Φ(–1.96 + –1.08)
= Φ(0.88) – Φ(-3.04) = 0.81057 – (0.00118) = 0.80939

c) $n=\dfrac{(z_{\alpha/2}+z_\beta)^2\sigma^2}{\delta^2}=\dfrac{(z_{0.025}+z_{0.10})^2\sigma^2}{(3.75-3)^2}=\dfrac{(1.96+1.29)^2(0.9)^2}{(0.75)^2}=15.21,\ \ n\cong 16$

9-44 a)

1) The parameter of interest is the true mean melting point, μ.
2) $H_0 : \mu = 155$
3) $H_1 : \mu \neq 155$
4) $z_0 = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
5) Reject H_0 if $z_0 < -z_{\alpha/2}$ where $\alpha = 0.01$ and $-z_{0.005} = -2.58$ or $z_0 > z_{\alpha/2}$ where $\alpha = 0.01$ and $z_{0.005} = 2.58$
6) $\bar{x} = 154.2$, $\sigma = 1.5$

$$z_0=\frac{154.2-155}{1.5/\sqrt{10}}=-1.69$$

7) Because –1.69 > -2.58, fail to reject the null hypothesis. There is not sufficient evidence to support the claim the mean melting point differs from 155 °F at α = 0.01.

b) P-value = 2P(Z <- 1.69) =2(0.045514) = 0.0910

c)

$$\beta=\Phi\left(z_{0.005}-\frac{\delta\sqrt{n}}{\sigma}\right)-\Phi\left(-z_{0.005}-\frac{\delta\sqrt{n}}{\sigma}\right)$$

$$=\Phi\left(2.58-\frac{(155-150)\sqrt{10}}{1.5}\right)-\Phi\left(-2.58-\frac{(155-150)\sqrt{10}}{1.5}\right)$$

= Φ(-7.96)- Φ(-13.12) = 0 – 0 = 0

d)

$$n = \frac{\left(z_{\alpha/2} + z_\beta\right)^2 \sigma^2}{\delta^2} = \frac{\left(z_{0.005} + z_{0.10}\right)^2 \sigma^2}{(150-155)^2} = \frac{(2.58+1.29)^2 (1.5)^2}{(5)^2} = 1.35,$$

n $\cong$ 2.

9-45 a)

1) The parameter of interest is the true mean battery life in hours, μ.
2) $H_0 : \mu = 40$
3) $H_1 : \mu > 40$
4) $z_0 = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$
5) Reject H_0 if $z_0 > z_\alpha$ where $\alpha = 0.05$ and $z_{0.05} = 1.65$
6) $\bar{x} = 40.5$, $\sigma = 1.25$

$$z_0 = \frac{40.5 - 40}{1.25/\sqrt{10}} = 1.26$$

7) Because 1.26 < 1.65, fail to reject H_0. There is not sufficient evidence to conclude that the mean battery life exceeds 40 at $\alpha = 0.05$.

b) P-value = $1 - \Phi(1.26) = 1 - 0.8962 = 0.1038$

c) $\beta = \Phi\left(z_{0.05} + \dfrac{40-42}{1.25/\sqrt{10}}\right)$ = $\Phi(1.65 + -5.06) = \Phi(-3.41) \cong 0.000325$

d) $n = \dfrac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{\delta^2} = \dfrac{\left(z_{0.05} + z_{0.10}\right)^2 \sigma^2}{(40-44)^2} = \dfrac{(1.65+1.29)^2 (1.25)^2}{(4)^2} = 0.844,\ n \cong 1$

e) 95% Confidence Interval

$$\bar{x} - z_{0.05}\sigma/\sqrt{n} \le \mu$$
$$40.5 - 1.65(1.25)/\sqrt{10} \le \mu$$
$$39.85 \le \mu$$

The lower bound of the 90 % confidence interval must be greater than 40 to verify that the true mean exceeds 40 hours.

9-46 a)

1) The parameter of interest is the true mean tensile strength, μ.
2) $H_0 : \mu = 3500$
3) $H_1 : \mu \ne 3500$
4) $z_0 = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$
5) Reject H_0 if $z_0 < -z_{\alpha/2}$ where $\alpha = 0.01$ and $-z_{0.005} = -2.58$ or $z_0 > z_{\alpha/2}$ where $\alpha = 0.01$ and $z_{0.005} = 2.58$
6) $\bar{x} = 3450$, $\sigma = 60$

$$z_0 = \frac{3450 - 3500}{60/\sqrt{12}} = -2.89$$

7) Because $-2.89 < -2.58$, reject the null hypothesis and conclude that the true mean tensile strength differs from 3500 at $\alpha = 0.01$.

b) Smallest level of significance = P-value = $2[1 - \Phi(2.89)] = 2[1 - .998074] = 0.004$. The smallest level of significance at which we reject the null hypothesis is 0.004.

c) $\delta = 3470 - 3500 = -30$

$$\beta = \Phi\left(z_{0.005} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{0.005} - \frac{\delta\sqrt{n}}{\sigma}\right)$$
$$= \Phi\left(2.58 - \frac{(3470-3500)\sqrt{12}}{60}\right) - \Phi\left(-2.58 - \frac{(3470-3500)\sqrt{12}}{60}\right)$$

= $\Phi(4.312)$- $\Phi(-0.848)$ = 1 – 0.1982 = 0.8018 so the power = 1- β = 0.20

d) $n = \dfrac{\left(z_{\alpha/2} + z_\beta\right)^2\sigma^2}{\delta^2} = \dfrac{\left(z_{0.005} + z_{0.20}\right)^2\sigma^2}{(3470-3500)^2} = \dfrac{(2.58+0.84)^2(60)^2}{(30)^2} = 46.79$

Therefore, the sample size that should be used equals 47.

e) 99% Confidence Interval

$$\bar{x} - z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$3450 - 2.58\left(\frac{60}{\sqrt{12}}\right) \le \mu \le 3450 + 2.58\left(\frac{60}{\sqrt{12}}\right)$$

$3405.313 \le \mu \le 3494.687$

With 99% confidence, the true mean tensile strength is between 3405.313 psi and 3494.687 psi. We can test the hypotheses that the true mean tensile strength is not equal to 3500 by noting that the value is not within the confidence interval. Hence, we reject the null hypothesis.

9-47 a)

1) The parameter of interest is the true mean speed, μ.
2) $H_0 : \mu = 100$
3) $H_1 : \mu < 100$
4) $z_0 = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
5) Reject H_0 if $z_0 < -z_\alpha$ where $\alpha = 0.05$ and $-z_{0.05} = -1.65$
6) $\bar{x} = 102.2$, $\sigma = 4$

$$z_0 = \frac{102.2-100}{4/\sqrt{8}} = 1.56$$

7) Because 1.56 > –1.65 fail to reject the null hypothesis. There is insufficient evidence to conclude that the true mean speed is less than 100 at $\alpha = 0.05$.

b) $z_0 = 1.56$, then P-value= $\Phi(z_0) \cong 0.94$

c) $\beta = 1 - \Phi\left(-z_{0.05} - \dfrac{(95-100)\sqrt{8}}{4}\right)$ = 1-Φ(-1.65 - –3.54) = 1-Φ(1.89) = 0.02938

Power = 1-β = 1-0.0294 = 0.9706

d) n = $\dfrac{\left(z_\alpha + z_\beta\right)^2\sigma^2}{\delta^2} = \dfrac{\left(z_{0.05} + z_{0.15}\right)^2\sigma^2}{(95-100)^2} = \dfrac{(1.65+1.03)^2(4)^2}{(5)^2} = 4.60,$ $n \cong 5$

e) 95% Confidence Interval

$$\mu \le \bar{x} + z_{0.05}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\mu \le 102.2 + 1.65\left(\frac{4}{\sqrt{8}}\right)$$

$$\mu \le 104.53$$

Because 100 is included in the CI, there is not sufficient evidence to reject the null hypothesis.

9-48 a) 1) The parameter of interest is the true mean hole diameter, μ.

2) $H_0 : \mu = 1.50$

3) $H_1 : \mu \ne 1.50$

4) $z_0 = \dfrac{\overline{x} - \mu}{\sigma / \sqrt{n}}$

5) Reject H_0 if $z_0 < -z_{\alpha/2}$ where $\alpha = 0.01$ and $-z_{0.005} = -2.58$ or $z_0 > z_{\alpha/2}$ where $z_{0.005} = 2.58$

6) $\bar{x} = 1.4975$, $\sigma = 0.01$

$$z_0 = \frac{1.4975 - 1.50}{0.01/\sqrt{25}} = -1.25$$

7) Because $-2.58 < -1.25 < 2.58$, fail to reject the null hypothesis. The mean hole diameter is not significantly different from 1.5 in. at $\alpha = 0.01$.

b) P-value=$2(1-\Phi(|Z_0|))=2(1-\Phi(1.25)) \cong 0.21$

c)

$$\beta = \Phi\left(z_{0.005} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{0.005} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

$$= \Phi\left(2.58 - \frac{(1.495-1.5)\sqrt{25}}{0.01}\right) - \Phi\left(-2.58 - \frac{(1.495-1.5)\sqrt{25}}{0.01}\right)$$

$= \Phi(5.08) - \Phi(-0.08) = 1 - 0.46812 = 0.53188$

Power = $1-\beta = 0.46812$.

d) Set $\beta = 1 - 0.90 = 0.10$

$$n = \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{\delta^2} = \frac{(z_{0.005} + z_{0.10})^2\sigma^2}{(1.495 - 1.50)^2} \cong \frac{(2.58 + 1.29)^2(0.01)^2}{(-0.005)^2} = 59.908,$$

$n \cong 60$

e) 99% confidence Interval

For $\alpha = 0.01$, $z_{\alpha/2} = z_{0.005} = 2.58$

$$\bar{x} - z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + z_{0.005}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$1.4975 - 2.58\left(\frac{0.01}{\sqrt{25}}\right) \le \mu \le 1.4975 + 2.58\left(\frac{0.01}{\sqrt{25}}\right)$$

$1.4923 \le \mu \le 1.5027$

The confidence interval constructed contains the value 1.5. Therefore, there is not strong evidence that true mean hole diameter differs from 1.5 in. using a 99% level of confidence. Because a two-sided 99% confidence interval is equivalent to a two-sided hypothesis test at $\alpha = 0.01$, the conclusions necessarily must be consistent.

9-49 a)

1) The parameter of interest is the true average battery life, μ.

2) $H_0 : \mu = 4$

3) $H_1 : \mu > 4$

4) $z_0 = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

5) Reject H_0 if $z_0 > z_\alpha$ where $\alpha = 0.05$ and $z_{0.05} = 1.65$

6) $\bar{x} = 4.05$, $\sigma = 0.2$

$$z_0 = \frac{4.05 - 4}{0.2/\sqrt{50}} = 1.77$$

7) Because 1.77 > 1.65, reject the null hypothesis. Conclude that the true average battery life exceeds 4 hours at α = 0.05.

b) P-value=1- $\Phi(Z_0)$ =1- $\Phi(1.77) \cong 0.04$

c) $\beta = \Phi\left(z_{0.05} - \dfrac{(4.5-4)\sqrt{50}}{0.2} \right)$ = $\Phi(1.65 – 17.68) = \Phi(-16.03) = 0$

Power = 1-β = 1-0 = 1

d) n = $\dfrac{(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2} = \dfrac{(z_{0.05} + z_{0.1})^2 \sigma^2}{(4.5-4)^2} = \dfrac{(1.65+1.29)^2 (0.2)^2}{(0.5)^2} = 1.38,$

n $\cong$ 2

e) 95% Confidence Interval

$$\bar{x} - z_{0.05}\left(\frac{\sigma}{\sqrt{n}} \right) \le \mu$$

$$4.05 - 1.65\left(\frac{0.2}{\sqrt{50}} \right) \le \mu$$

$$4.003 \le \mu$$

Because the lower limit of the CI is greater than 4, we conclude that average life is greater than 4 hours at α = 0.05.

9-50

a) The alternative hypothesis is one-sided since we are testing whether the mean gestation period is less than 280 days for patients at risk.

b)

1) The parameter of interest is the true mean gestation period, μ .

2) H_0 : μ = 280

3) H_1 : μ < 280

4) $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

5) Reject H_0 if $z_0 < -z_\alpha$ where α = 0.05 and $z_{0.05}$ = 1.65

6) $\bar{x} = 274.3$, $\sigma = 9$

$$z_0 = \frac{274.3 - 280}{9/\sqrt{70}} = -5.299$$

7) Because -5.299 < -1.65, reject H_0. The gestation period is significantly less than 280 at α = 0.05.

c) $z_0 = -5.299$ and the P-value = $\Phi(z_0) \cong 0$

9-51

a) The alternative hypothesis should be one-sided because the scientists are interested in high adhesion.
b)
1) The parameter of interest is the true mean adhesion, μ .
2) H_0 : $\mu = 2.5$
3) H_1 : $\mu > 2.5$
4) $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
5) Reject H_0 if $z_0 > z_\alpha$ where α = 0.05 and $z_{0.05} = 1.65$
6) $\bar{x} = 3.372$, $\sigma = 0.66$

$$z_0 = \frac{3.372 - 2.5}{0.66/\sqrt{5}} = 2.95$$

7) Because 2.95 > 1.65, reject H_0. The true mean adhesion is greater than 2.5 at α = 0.05.
c) P-value = $1 - \Phi(2.95) = 1 - 0.998 = 0.002$

Section 9-3

9-52 a) α = 0.01, n=20, the critical values are ± 2.861
b) α = 0.05, n=12, the critical values are ± 2.201
c) α = 0.1, n=15, the critical values are ± 1.761

9-53 a) α = 0.01, n = 20, the critical value = 2.539
b) α = 0.05, n = 12, the critical value = 1.796
c) α = 0.1, n = 15, the critical value = 1.345

9-54 a) α = 0.01, n = 20, the critical value = -2.539
b) α = 0.05, n = 12, the critical value = -1.796
c) α = 0.1, n = 15, the critical value = -1.345

9-55 a) $2*0.025 \le p \le 2*0.05$ then $0.05 \le p \le 0.1$
b) $2*0.025 \le p \le 2*0.05$ then $0.05 \le p \le 0.1$
c) $2*0.25 \le p \le 2*0.4$ then $0.5 \le p \le 0.8$

9-56 a) $0.025 \le p \le 0.05$
b) $1 - 0.05 \le p \le 1 - 0.025$ then $0.95 \le p \le 0.975$
c) $0.25 \le p \le 0.4$

9-57 a) $1 - 0.05 \le p \le 1 - 0.025$ then $0.95 \le p \le 0.975$
b) $0.025 \le p \le 0.05$
c) $1 - 0.4 \le p \le 1 - 0.25$ then $0.6 \le p \le 0.75$

9-58 a) SE Mean $= \dfrac{S}{\sqrt{N}} = \dfrac{0.717}{\sqrt{20}} = 0.1603$

$$t_0 = \frac{92.379 - 91}{0.717/\sqrt{20}} = 8.6012$$

$t_0 = 8.6012$ with df = 20 – 1 = 19, so the P-value < 0.0005. Because the P-value < α = 0.05 we reject the null hypothesis that μ = 91at the 0.05 level of significance.

95% lower CI of the mean is $\bar{x} - t_{0.05,19}\dfrac{S}{\sqrt{n}} \le \mu$

$$92.379 - (1.729)\frac{0.717}{\sqrt{20}} \le \mu$$
$$92.1018 \le \mu$$

b) A one-sided test because the alternative hypothesis is μ > 91.

c) If the alternative hypothesis is changed to μ > 90, then $t_0 = \dfrac{92.379 - 90}{0.717/\sqrt{20}} = 14.8385$

$t_0 = 14.8385$ with df = 20 – 1 = 19, so the P-value < 0.0005. The P-value < α = 0.05 and we reject the null hypothesis at the 0.05 level of significance.

9-59 a) degrees of freedom = n - 1 = 10 - 1 = 9

b) SE Mean $= \dfrac{s}{\sqrt{N}} = \dfrac{s}{\sqrt{10}} = 0.296$, then s = 0.9360.

$$t_0 = \frac{12.564 - 12}{0.296} = 1.905$$

$t_0 = 1.905$ with df = 10 – 1 = 9.

The P-value falls between two values: 1.833 (for α = 0.05) and 2.262 (for α = 0.025), so 0.05 = 2(0.025) < P-value < 2(0.05) = 0.1. The P-value > α = 0.05, so we fail to reject the null hypothesis at the 0.05 level of significance.

c) A two-sided test because the alternative hypothesis is μ ≠ 12.

d) 95% two-sided CI

$$\bar{x} - t_{0.025,9}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,9}\left(\frac{s}{\sqrt{n}}\right)$$
$$12.564 - 2.262\left(\frac{0.9360}{\sqrt{10}}\right) \le \mu \le 12.564 + 2.262\left(\frac{0.9360}{\sqrt{10}}\right)$$
$$11.8945 \le \mu \le 13.2335$$

e) Suppose that the alternative hypothesis is changed to μ > 12. Because $t_0 = 1.905 > t_{0.05,9} = 1.833$ we reject the null hypothesis at the 0.05 level of significance.

f) Reject the null hypothesis that μ = 11.5 versus the alternative hypothesis (μ ≠ 11.5) at the 0.05 level of significance because the μ = 11.5 is not included in the 95% two-sided CI on the mean.

9-60 a) degrees of freedom = N – 1 = 16 – 1 = 15

b) SE Mean $= \dfrac{S}{\sqrt{N}} = \dfrac{1.783}{\sqrt{16}} = 0.4458$

$$t_0 = \frac{35.274 - 34}{1.783/\sqrt{16}} = 2.8581$$

c) P-value = 2P(t > 2.8581) = 0.012. We reject the null hypothesis if the P-value < α. Thus, we can reject the null hypothesis at significance levels greater than 0.012.

d) If the alternative hypothesis is changed to the one-sided alternative μ > 34, the P-value = 0.5(0.12) = 0.006.

e) If the null hypothesis is changed to μ = 34.5 versus the alternative hypothesis (μ ≠ 34.5) the t statistic is reduced. In particular, $t_0 = \dfrac{35.274 - 34.5}{1.783/\sqrt{16}} = 1.7364$ and $t_{0.025,15} = 2.131$.

Because $t_0 = 1.7364 < t_{0.025,15}$, we fail to reject the null hypothesis at the 0.05 level of significance.

9-61 a)

1) The parameter of interest is the true mean of body weight, μ.
2) H_0: μ = 300
3) H_1: μ ≠ 300
4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$
5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$ where α = 0.05 and $t_{\alpha/2,n-1}$ = 2.056 for n = 27
6) $\bar{x} = 325.496$, *s* = 198.786, n = 27

$$t_0 = \frac{325.496 - 300}{198.786/\sqrt{27}} = 0.6665$$

7) Because 0.6665 < 2.056 we fail to reject the null hypothesis. There is not sufficient evidence to conclude that the true mean body weight differs from 300 at α = 0.05. We have 2(0.25) < P-value < 2(0.4). That is, 0.5 < P-value < 0.8

b) We reject the null hypothesis if P-value < α. The P-value = 2(0.2554) = 0.5108. Therefore, the smallest level of significance at which we can reject the null hypothesis is approximately 0.51.

c) 95% two sided confidence interval

$$\bar{x} - t_{0.025,26}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,26}\left(\frac{s}{\sqrt{n}}\right)$$

$$325.496 - 2.056\left(\frac{198.786}{\sqrt{27}}\right) \le \mu \le 325.496 + 2.056\left(\frac{198.786}{\sqrt{27}}\right)$$

$$246.8409 \le \mu \le 404.1511$$

We fail to reject the null hypothesis because the hypothesized value of 300 is included within the confidence interval.

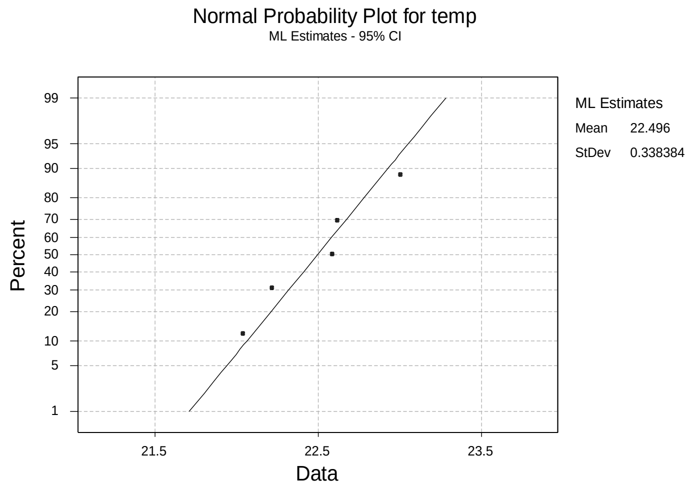
9-62 a)

1) The parameter of interest is the true mean interior temperature life, μ.
2) H_0 : μ = 22.5
3) H_1 : μ ≠ 22.5
4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$
5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$ where α = 0.05 and $t_{\alpha/2,n-1}$ = 2.776 for n = 5
6) $\bar{x} = 22.496$, *s* = 0.378, *n* = 5

$$t_0 = \frac{22.496 - 22.5}{0.378/\sqrt{5}} = -0.00237$$

7) Because –0.00237 > –2.776, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that the true mean interior temperature differs from 22.5 °C at α = 0.05.
Also, 2(0.4) < *P*-value < 2(0.5). That is, 0.8 < *P*-value < 1.0.

b) The points on the normal probability plot fall along a line. Therefore, the normality assumption is reasonable.



c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|22.75 - 22.5|}{0.378} = 0.66$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, $d = 0.66$, and n = 5, we obtain $\beta \cong 0.8$ and power of $1–0.8 = 0.2$

d) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|22.75 - 22.5|}{0.378} = 0.66$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, $d = 0.66$, and $\beta \cong 0.1$ (Power=0.9), n = 40

e) 95% two sided confidence interval

$$\bar{x} - t_{0.025,4}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,4}\left(\frac{s}{\sqrt{n}}\right)$$

$$22.496 - 2.776\left(\frac{0.378}{\sqrt{5}}\right) \le \mu \le 22.496 + 2.776\left(\frac{0.378}{\sqrt{5}}\right)$$

$$22.027 \le \mu \le 22.965$$

We cannot conclude that the mean interior temperature differs from 22.5 at $\alpha = 0.05$ because the value is included in the confidence interval.

9-63 a)

1) The parameter of interest is the true mean female body temperature, μ.

2) $H_0 : \mu = 98.6$

3) $H_1 : \mu \neq 98.6$

4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$ where $\alpha = 0.05$ and $t_{\alpha/2,n-1} = 2.064$ for n = 25

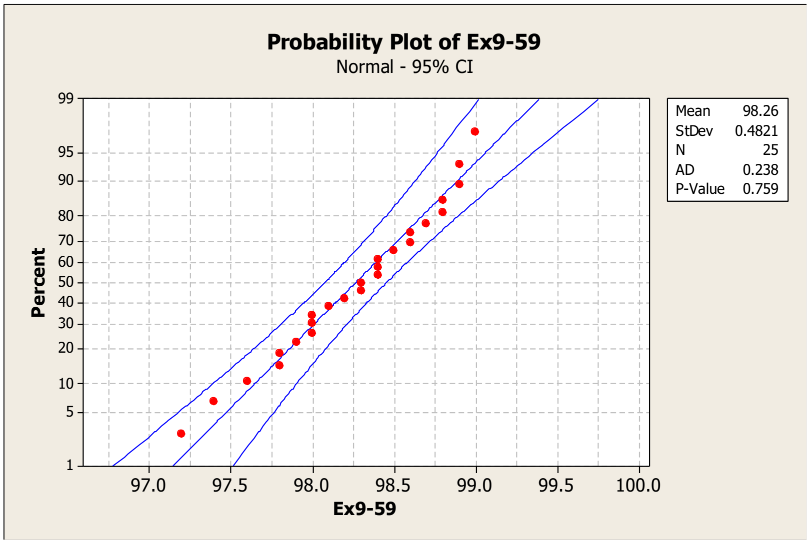
6) $\bar{x} = 98.264$, s = 0.4821, n = 25

$$t_0 = \frac{98.264 - 98.6}{0.4821/\sqrt{25}} = -3.48$$

7) Because 3.48 > 2.064, reject the null hypothesis. Conclude that the true mean female body temperature differs from 98.6 °F at $\alpha = 0.05$.

P-value = 2(0.001) = 0.002

b) The data on the normal probability plot falls along a line. The normality assumption is reasonable.



c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|98 - 98.6|}{0.4821} = 1.24$

Using the OC curve, Chart VII*e* for $\alpha = 0.05$, d = 1.24, and n = 25, obtain $\beta \cong 0$ and power of $1 - 0 \cong 1$

d) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|98.2 - 98.6|}{0.4821} = 0.83$

Using the OC curve, Chart VII*e* for $\alpha = 0.05$, d = 0.83, and $\beta \cong 0.1$ (Power=0.9), n =20

e) 95% two sided confidence interval

$$\bar{x} - t_{0.025,24}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,24}\left(\frac{s}{\sqrt{n}}\right)$$

$$98.264 - 2.064\left(\frac{0.4821}{\sqrt{25}}\right) \le \mu \le 98.264 + 2.064\left(\frac{0.4821}{\sqrt{25}}\right)$$

$$98.065 \le \mu \le 98.463$$

We conclude that the mean female body temperature differs from 98.6 at $\alpha = 0.05$ because the value is not included inside the confidence interval.

9-64 a)

1) The parameter of interest is the true mean rainfall, μ.

2) $H_0 : \mu = 25$

3) $H_1 : \mu > 25$

4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

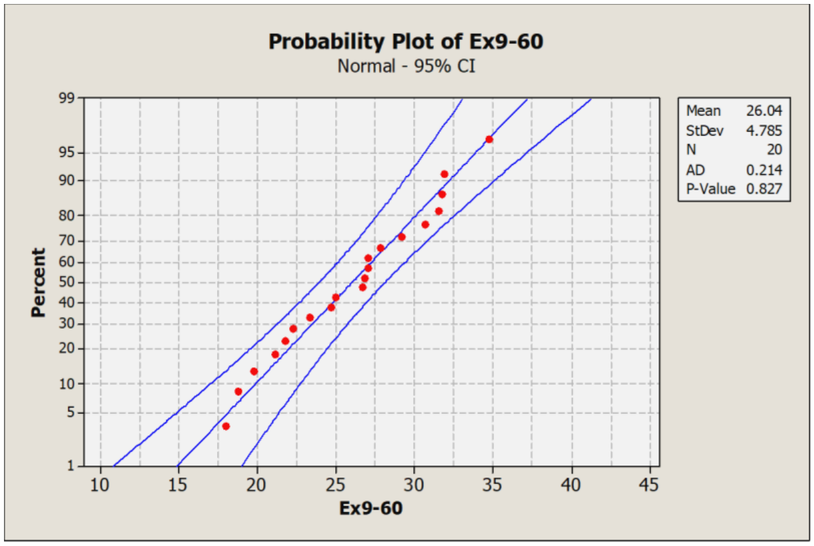
5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.01$ and $t_{0.01,19} = 2.539$ for n = 20

6) $\bar{x} = 26.04$ s = 4.78 n = 20

$$t_0 = \frac{26.04 - 25}{4.78/\sqrt{20}} = 0.97$$

7) Because 0.97 < 2.539 fail to reject the null hypothesis. There is insufficient evidence to conclude that the true mean rainfall is greater than 25 acre-feet at $\alpha = 0.01$. The 0.10 < P-value < 0.25.

b) The data on the normal probability plot falls along a line. Therefore, the normality assumption is reasonable.



c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|27 - 25|}{4.78} = 0.42$

Using the OC curve, Chart VII h) for $\alpha = 0.01$, d = 0.42, and n = 20, obtain $\beta \cong 0.7$ and power of 1 – 0.7 = 0.3.

d) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|27.5 - 25|}{4.78} = 0.52$

Using the OC curve, Chart VII h) for $\alpha = 0.05$, d = 0.42, and $\beta \cong 0.1$ (Power=0.9), n = 75

e) 99% lower confidence bound on the mean diameter

$$\bar{x} - t_{0.01,19}\left(\frac{s}{\sqrt{n}}\right) \le \mu$$

$$26.04 - 2.539\left(\frac{4.78}{\sqrt{20}}\right) \le \mu$$

$$23.326 \le \mu$$

Because the lower limit of the CI is less than 25 there is insufficient evidence to conclude that the true mean rainfall is greater than 25 acre-feet at $\alpha = 0.01$.

9-65 a)

1) The parameter of interest is the true mean sodium content, μ.

2) $H_0 : \mu = 130$

3) $H_1 : \mu \ne 130$

4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$ where $\alpha = 0.05$ and $t_{\alpha/2,n-1} = 2.093$ for n = 20

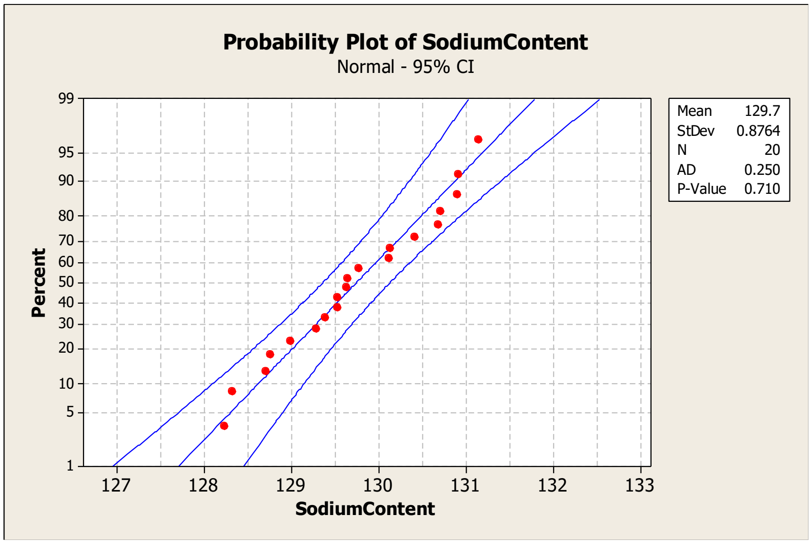
6) $\bar{x} = 129.747$, s = 0.876 n = 20

$$t_0 = \frac{129.747 - 130}{0.876/\sqrt{20}} = -1.291$$

7) Because 1.291< 2.093 we fail to reject the null hypothesis. There is not sufficient evidence that the true mean sodium content is different from 130mg at $\alpha = 0.05$.

From the *t* table (Table V) the t_0 value is between the values of 0.1 and 0.25 with 19 degrees of freedom. Therefore, 2(0.1) < P-value < 2(0.25) or 0.2 < P-value < 0.5.

b) The assumption of normality appears to be reasonable.



c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|130.5 - 130|}{0.876} = 0.571$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, d = 0.57, and n = 20, we obtain $\beta \cong 0.3$ and the power of 1 – 0.30 = 0.70

d) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|130.1 - 130|}{0.876} = 0.114$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, d = 0.11, and $\beta \cong 0.25$ (Power = 0.75), the sample sizes do not extend to the point $d = 0.114$ and $\beta = 0.25$. We can conclude that n > 100

e) 95% two sided confidence interval

$$\bar{x} - t_{0.025,29}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,29}\left(\frac{s}{\sqrt{n}}\right)$$

$$129.747 - 2.093\left(\frac{0.876}{\sqrt{20}}\right) \le \mu \le 129.747 + 2.093\left(\frac{0.876}{\sqrt{20}}\right)$$

$$129.337 \le \mu \le 130.157$$

There is no evidence that the mean differs from 130 because that value is w the confidence interval. The result is the same as part (a).

9-66

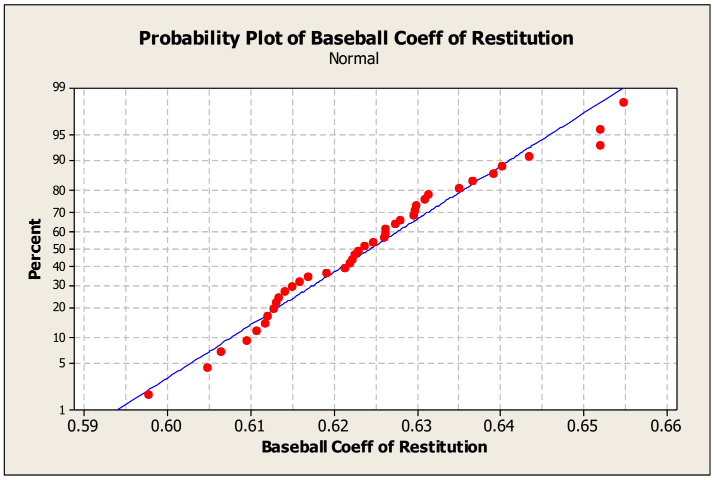
a) In order to use t statistics in hypothesis testing, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true mean coefficient of restitution, μ.
2) $H_0 : \mu = 0.635$
3) $H_1 : \mu > 0.635$
4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$
5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.05$ and $t_{0.05,39} = 1.685$ for n = 40
6) $\bar{x} = 0.624$ s = 0.013 n = 40

$$t_0 = \frac{0.624 - 0.635}{0.013/\sqrt{40}} = -5.35$$

7) Because –5.25 < 1.685 fail to reject the null hypothesis. There is not sufficient evidence to conclude that the true mean coefficient of restitution is greater than 0.635 at $\alpha = 0.05$.

The area to right of -5.35 under the *t* distribution is greater than 0.9995 from Table V.

b) From the normal probability plot, the normality assumption seems reasonable:



c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|0.64 - 0.635|}{0.013} = 0.38$

Using the OC curve, Chart VII g) for $\alpha = 0.05$, d = 0.38, and n = 40, obtain $\beta \cong 0.25$ and power of 1 – 0.25 = 0.75.

d) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|0.638 - 0.635|}{0.013} = 0.23$

Using the OC curve, Chart VII g) for $\alpha = 0.05$, d = 0.23, and $\beta \cong 0.25$ (Power = 0.75), n = 40

e) 95% lower confidence bound is $\bar{x} - t_{\alpha,n-1}\left(\dfrac{s}{\sqrt{n}}\right) = 0.6205$

Because 0.635 > 0.6205, we fail to reject the null hypothesis.

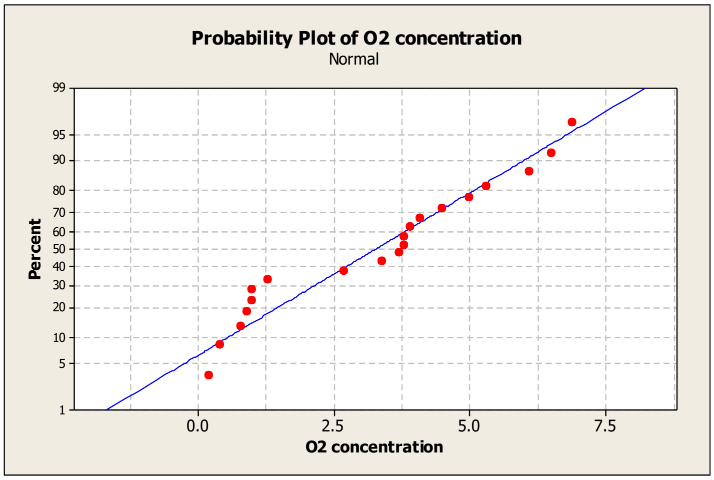
9-67

a) In order to use t statistics in hypothesis testing, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true mean oxygen concentration, μ.
2) $H_0 : \mu = 4$
3) $H_1 : \mu \neq 4$
4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$
5) Reject H_0 if $|t_0| > t_{\alpha/2,\, n-1}$ where $\alpha = 0.01$ and $t_{0.005,\,19} = 2.861$ for n = 20
6) $\bar{x} = 3.265$, s = 2.127, n = 20

$$t_0 = \frac{3.265 - 4}{2.127/\sqrt{20}} = -1.55$$

7) Because -2.861< -1.55, fail to reject the null hypothesis. There is insufficient evidence to conclude that the true mean oxygen differs from 4 at $\alpha = 0.01$. Also 2(0.05) < *P*-value < 2(0.10). Therefore 0.10 < *P*-value < 0.20

b) From the normal probability plot, the normality assumption seems reasonable.



c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|3-4|}{2.127} = 0.47$

Using the OC curve, Chart VII f) for $\alpha = 0.01$, d = 0.47, and n = 20, we obtain $\beta \cong 0.70$ and power of 1–0.70 = 0.30.

d) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|2.5-4|}{2.127} = 0.71$

Using the OC curve, Chart VII f) for $\alpha = 0.01$, d = 0.71, and $\beta \cong 0.10$ (Power=0.90), $n = 40$.

e) The 95% confidence interval is

$$\bar{x} - t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right) = 1.9 \le \mu \le 4.62$$

Because 4 is within the confidence interval, we fail to reject the null hypothesis.

9-68 a)

1) The parameter of interest is the true mean sodium content, μ.

2) $H_0 : \mu = 300$

3) $H_1 : \mu > 300$

4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.05$ and $t_{\alpha,n-1} = 1.943$ for n = 7

6) $\bar{x} = 315$, s = 16 n=7

$$t_0 = \frac{315 - 300}{16/\sqrt{7}} = 2.48$$

7) Because 2.48 > 1.943 reject the null hypothesis and conclude that the leg strength exceeds 300 watts at $\alpha = 0.05$.

The P-value is between 0.01 and 0.025

b) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|305-300|}{16} = 0.3125$

Using the OC curve, Chart VII g) for $\alpha = 0.05$, d = 0.3125, and n = 7, $\beta \cong 0.9$ and power = 1–0.9 = 0.1.

c) If $1 - \beta > 0.9$ then $\beta < 0.1$ and n is approximately 100

d) Lower confidence bound is $\bar{x} - t_{\alpha,n-1}\left(\dfrac{s}{\sqrt{n}}\right) = 303.2 < \mu$

Because 300 is not include in the interval, reject the null hypothesis

9-69 a)

1) The parameter of interest is the true mean tire life, μ.
2) $H_0 : \mu = 60000$
3) $H_1 : \mu > 60000$
4) $t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$
5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.05$ and $t_{0.05,15} = 1.753$ for n = 16
6) $n = 16 \quad \bar{x} = 60{,}139.7 \quad s = 3645.94$

$$t_0 = \frac{60139.7-60000}{3645.94/\sqrt{16}} = 0.15$$

7) Because 0.15 < 1.753 fail to reject the null hypothesis. There is insufficient evidence to conclude that the true mean tire life is greater than 60,000 kilometers at $\alpha = 0.05$. The P-value > 0.40.

b) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu-\mu_0|}{\sigma} = \dfrac{|61000-60000|}{3645.94} = 0.27$

Using the OC curve, Chart VII g) for $\alpha = 0.05$, d = 0.27, and $\beta \cong 0.1$ (Power = 0.9), n = 4.
Yes, the sample size of 16 was sufficient.

9-70 In order to use t statistics in hypothesis testing, we need to assume that the underlying distribution is normal.

1) The parameter of interest is the true mean impact strength, μ.
2) $H_0 : \mu = 1.0$
3) $H_1 : \mu > 1.0$
4) $t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$
5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.05$ and $t_{0.05,19} = 1.729$ for n = 20
6) $\bar{x} = 1.25 \quad s = 0.25 \quad n = 20$

$$t_0 = \frac{1.25-1.0}{0.25/\sqrt{20}} = 4.47$$

7) Because 4.47 > 1.729 reject the null hypothesis. There is sufficient evidence to conclude that the true mean impact strength is greater than 1.0 ft-lb/in at $\alpha = 0.05$. The P-value < 0.0005

9-71 In order to use a t statistic in hypothesis testing, we need to assume that the underlying distribution is normal.

1) The parameter of interest is the true mean current, μ.
2) $H_0 : \mu = 300$
3) $H_1 : \mu > 300$
4) $t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$
5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.05$ and $t_{0.05,9} = 1.833$ for n = 10
6) $n = 10 \quad \bar{x} = 317.2 \quad s = 15.7$

$$t_0 = \frac{317.2-300}{15.7/\sqrt{10}} = 3.46$$

7) Because 3.46 > 1.833, reject the null hypothesis. There is sufficient evidence to indicate that the true mean current is greater than 300 microamps at $\alpha = 0.05$. The 0.0025 <*P*-value < 0.005

9-72 a)

1) The parameter of interest is the true mean height of female engineering students, μ.
2) $H_0 : \mu = 65$
3) $H_1 : \mu > 65$

4) $t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

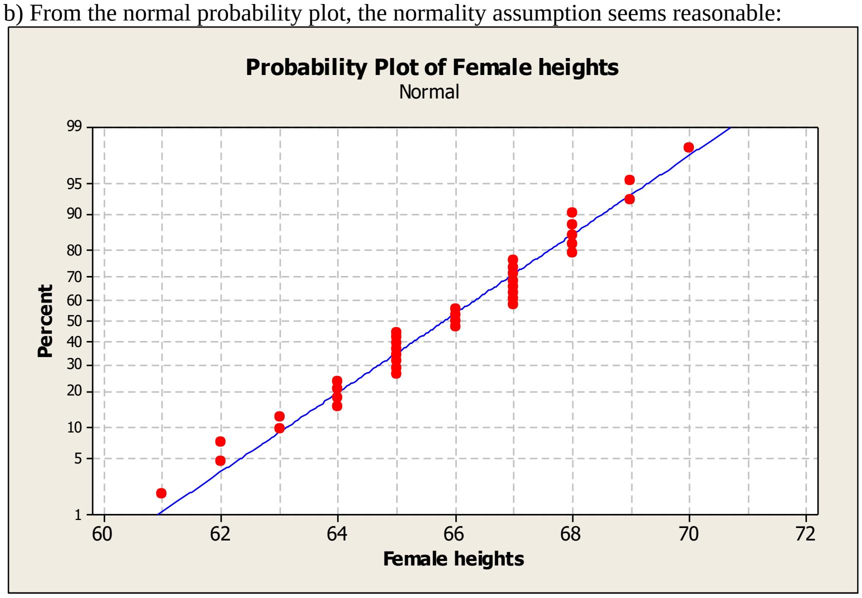
5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.05$ and $t_{0.05,36} = 1.68$ for n = 37

6) $\bar{x} = 65.811$ inches $s = 2.106$ inches n = 37

$$t_0 = \frac{65.811-65}{2.11/\sqrt{37}} = 2.34$$

7) Because 2.34 > 1.68 reject the null hypothesis. There is sufficient evidence to conclude that the true mean height of female engineering students is greater than 65 at $\alpha = 0.05$. We obtain $0.01 < P\text{-value} < 0.025$.

b) From the normal probability plot, the normality assumption seems reasonable:



c) $d = \dfrac{|68-65|}{2.11} = 1.42$, n=37

From the OC Chart VII g) for $\alpha = 0.05$, we obtain $\beta \cong 0$. Therefore, the power $\cong 1$.

d) $d = \dfrac{|66-65|}{2.11} = 0.47$

From the OC Chart VIIg) for $\alpha = 0.05$ and $\beta \cong 0.2$ (power = 0.8). Therefore, $n \cong 30$.

9-73 a) In order to use t statistics in hypothesis testing, we need to assume that the underlying distribution is normal.

1) The parameter of interest is the true mean distance, μ.

2) $H_0 : \mu = 280$

3) $H_1 : \mu > 280$

4) $t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $\alpha = 0.05$ and $t_{0.05,99} = 1.6604$ for n = 100

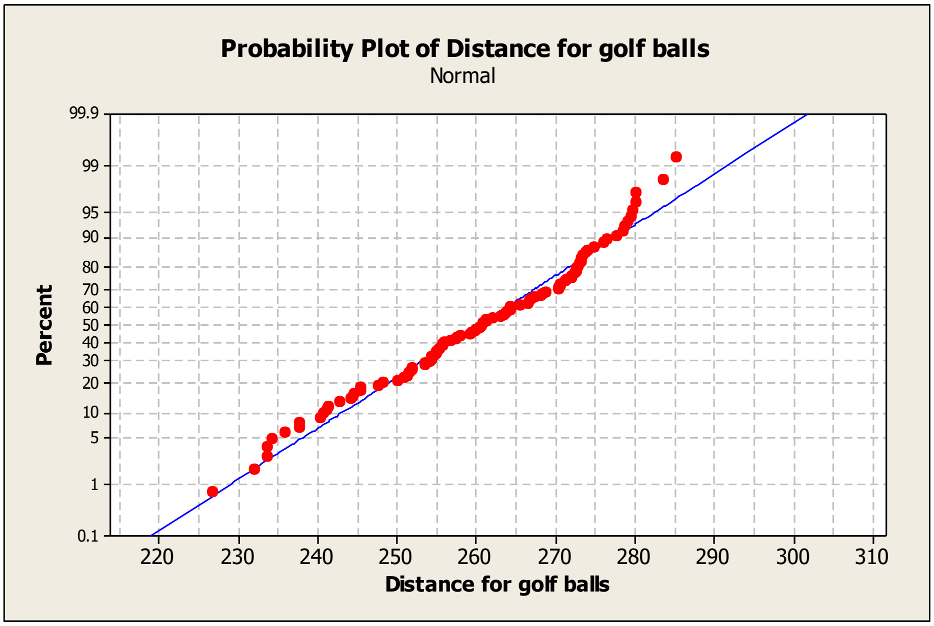
6) $\bar{x} = 260.3$ s = 13.41 n = 100

$$t_0 = \frac{260.3-280}{13.41/\sqrt{100}} = -14.69$$

7) Because –14.69 < 1.6604 fail to reject the null hypothesis. There is insufficient evidence to conclude that the true mean distance is greater than 280 at $\alpha = 0.05$.

From Table V, the t_0 value in absolute value is greater than the value corresponding to 0.0005. Therefore, P-value > 0.9995.

b) From the normal probability plot, the normality assumption seems reasonable:



c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|290 - 280|}{13.41} = 0.75$

Using the OC curve, Chart VII g) for $\alpha = 0.05$, d = 0.75, and n = 100, obtain $\beta \cong 0$ and power ≈ 1

d) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|290 - 280|}{13.41} = 0.75$

Using the OC curve, Chart VII g) for $\alpha = 0.05$, d = 0.75, and $\beta \cong 0.20$ (Power = 0.80), n =15

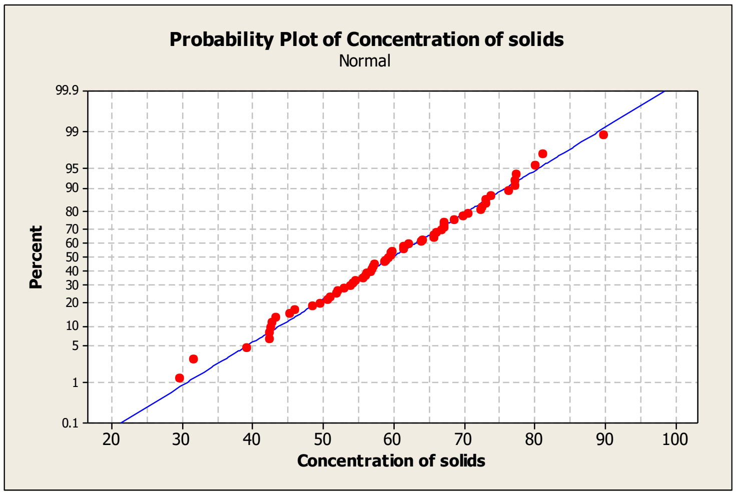
9-74 a) In order to use t statistics in hypothesis testing, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true mean concentration of suspended solids, μ.
2) $H_0 : \mu = 55$
3) $H_1 : \mu \neq 55$
4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$
5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$ where $\alpha = 0.05$ and $t_{0.025,59} = 2.000$ for n = 60
6) $\bar{x} = 59.87$ s = 12.50 n = 60

$$t_0 = \frac{59.87 - 55}{12.50/\sqrt{60}} = 3.018$$

7) Because 3.018 > 2.000, reject the null hypothesis. There is sufficient evidence to conclude that the true mean concentration of suspended solids is not equal to 55 at $\alpha = 0.05$.

From Table V the t_0 value is between the values of 0.001 and 0.0025 with 59 degrees of freedom. Therefore, 2(0.001) < P-value < 2(0.0025) and 0.002 < P-value < 0.005. Computer software generates a P-value = 0.0038.

b) The data tend to fall along a line. The normality assumption seems reasonable.



c) $d = \dfrac{|50-55|}{12.50} = 0.4$, n = 60 so, from the OC Chart VII e) for $\alpha = 0.05$, d= 0.4 and n = 60 obtain $\beta \cong 0.2$.
Therefore, the power = 1 - 0.2 = 0.8

d) From the same OC chart, and for the specified power, we would need approximately 75 observations.

$$d = \frac{|50-55|}{12.50} = 0.4$$

Using the OC Chart VII e) for $\alpha = 0.05$, d = 0.4, and $\beta \cong 0.10$ so that the power = 0.90, n = 75

9-75

a)

$H_0 : \mu = 98.2$

$H_1 : \mu \neq 98.2$

b)

1) The parameter of interest is the true mean oral temperature, μ

2) $H_0 : \mu = 98.2$

3) $H_1 : \mu \neq 98.2$

4) $t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

5) Reject H_0 if $|t_0| > t_{\alpha/2,\,n-1}$ where $\alpha = 0.05$ and $t_{0.025,\,51} = 2.008$ for n = 52

6) $\bar{x} = 98.285$, s = 0.625, n = 52

$$t_0 = \frac{98.285-98.2}{0.625/\sqrt{52}} = 0.981$$

7) As the sample statistic is less than the table value, i.e. 0.981 < 2.008, we fail to reject the null hypothesis at $\alpha = 0.05$. There is insufficient evidence to conclude that the true mean oral temperature differs from 98.2 at $\alpha = 0.05$.

c) 95% two sided confidence interval

$$\bar{x} - t_{0.025,51}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,51}\left(\frac{s}{\sqrt{n}}\right)$$

$$98.285 - 2.008\left(\frac{0.625}{\sqrt{52}}\right) \le \mu \le 98.285 + 2.008\left(\frac{0.625}{\sqrt{52}}\right)$$

$$98.111 \le \mu \le 98.459$$

There is no evidence that the mean differs from 98.2 because that value is within the confidence interval. The result is the same as part (b).

9-76

a) 95% two sided confidence interval

$$\bar{x} - t_{0.025,99}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,99}\left(\frac{s}{\sqrt{n}}\right)$$

$$299{,}852.4 - 1.984\left(\frac{79.01}{\sqrt{100}}\right) \le \mu \le 299{,}852.4 + 1.984\left(\frac{79.01}{\sqrt{100}}\right)$$

$$299{,}836.7244 \le \mu \le 299{,}868.0756$$

b) Michelson's measurements do not seem to be accurate because the true value of 299,734.5 is not included in the confidence interval at $\alpha = 0.05$.

Section 9-4

9-77 a) α = 0.01, n = 20, from Table V we find the following critical values 6.84 and 38.58
b) α = 0.05, n = 12, from Table V we find the following critical values 3.82 and 21.92
c) α = 0.10, n = 15, from Table V we find the following critical values 6.57 and 23.68

9-78 a) α = 0.01, n = 20, from Table V we find $\chi^2_{\alpha,n-1} = 36.19$

b) α = 0.05, n = 12, from Table V we find $\chi^2_{\alpha,n-1} = 19.68$

c) α = 0.10, n = 15, from Table V we find $\chi^2_{\alpha,n-1} = 21.06$

9-79 a) α = 0.01, n = 20, from Table V we find $\chi^2_{1-\alpha,n-1} = 7.63$

b) α = 0.05, n = 12, from Table V we find $\chi^2_{1-\alpha,n-1} = 4.57$

c) α = 0.10, n = 15, from Table V we find $\chi^2_{1-\alpha,n-1} = 7.79$

9-80 a) 2(0.1) < *P*-value < 2(0.5), then 0.2 < *P*-value < 1
b) 2(0.1) < *P*-value < 2(0.5), then 0.2 < *P*-value < 1
c) 2(0.05) < *P*-value < 2(0.1), then 0.1 < *P*-value < 0.2

9-81 a) 0.1 < 1-*P* < 0.5 then 0.5 < *P*-value < 0.9
b) 0.1 < 1-*P*< 0.5 then 0.5 < *P*-value < 0.9
c) 0.99 <1-*P* <0.995 then 0.005 < *P*-value < 0.01

9-82 a) 0.1 < *P*-value < 0.5
b) 0.1 < *P*-value < 0.5
c) 0.99 < *P*-value < 0.995

9-83

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true standard deviation of performance time σ. However, the solution can be found by performing a hypothesis test on σ^2.

2) H_0 : $\sigma^2 = 0.75^2$
3) H_1 : $\sigma^2 > 0.75^2$
4) $\chi^2_0 = \frac{(n-1)s^2}{\sigma^2}$

5) Reject H_0 if $\chi^2_0 > \chi^2_{\alpha,n-1}$ where α = 0.05 and $\chi^2_{0.05,16} = 26.30$
6) n = 17, s = 0.09

$$\chi^2_0 = \frac{(n-1)s^2}{\sigma^2} = \frac{16(0.09)^2}{.75^2} = 0.23$$

7) Because 0.23 < 26.30, fail to reject H_0. There is insufficient evidence to conclude that the true variance of performance time content exceeds 0.75^2 at α = 0.05. Because χ^2_0 = 0.23, the *P*-value > 0.995

b) The 95% one sided confidence interval given below includes the value 0.75. Therefore, we are not be able to conclude that the standard deviation is greater than 0.75.

$$\frac{16(.09)^2}{26.3} \le \sigma^2$$
$$0.07 \le \sigma$$

9-84

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.

1) The parameter of interest is the true measurement standard deviation σ. However, the answer can be found by performing a hypothesis test on σ^2.

2) H_0 : $\sigma^2 = .01^2$

3) H_1 : $\sigma^2 \neq .01^2$

4) $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma^2}$

5) Reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.975,14} = 5.63$ or $\chi_0^2 > \chi^2_{\alpha,2,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.025,14} = 26.12$ for n = 15

6) n = 15, s = 0.0083

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{14(.0083)^2}{.01^2} = 9.6446$$

7) Because 5.63 < 9.64 < 26.12 fail to reject H_0. 0.1 < *P*-value/2 < 0.5. Therefore, 0.2 < *P*-value < 1

b) The 95% confidence interval includes the value 0.01. Therefore, there is not enough evidence to reject the null hypothesis.

$$\frac{14(.0083)^2}{26.12} \leq \sigma^2 \leq \frac{14(.0083)^2}{5.63}$$
$$0.00607 \leq \sigma^2 \leq 0.013$$

9-85

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.

1) The parameter of interest is the true standard deviation of titanium percentage, σ. However, the solution can be found by performing a hypothesis test on σ^2.

2) H_0 : $\sigma^2 = (0.25)^2$

3) H_1 : $\sigma^2 \neq (0.25)^2$

4) $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma^2}$

5) Reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.995,50} = 32.36$ or $\chi_0^2 > \chi^2_{\alpha,2,n-1}$ where $\alpha = 0.05$ and

$\chi^2_{0.005,50} = 71.42$ for n = 51

6) n = 51, s = 0.37

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{50(0.37)^2}{(0.25)^2} = 109.52$$

7) Because 109.52 > 71.42 reject H_0. The standard deviation of titanium percentage is significantly different from 0.25 at $\alpha = 0.01$. P-value/2 < 0.005, then P-value < 0.01

b) 95% confidence interval for σ:

First find the confidence interval for σ^2 :

For $\alpha = 0.05$ and n = 51, $\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,50} = 71.42$ and $\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,50} = 32.36$

$$\frac{50(0.37)^2}{71.42} \leq \sigma^2 \leq \frac{50(0.37)^2}{32.36}$$
$$0.096 \leq \sigma^2 \leq 0.2115$$

Taking the square root of the endpoints of this interval we obtain, $0.31 < \sigma < 0.46$

Because 0.25 falls below the lower confidence bound we conclude that the population standard deviation is not equal to 0.25.

9-86

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true standard deviation of Izod impact strength, σ. However, the solution can be found by performing a hypothesis test on σ^2.
2) $H_0: \sigma^2 = (0.10)^2$
3) $H_1: \sigma^2 \neq (0.10)^2$
4) $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$
5) Reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$ where $\alpha = 0.01$ and $\chi^2_{0.995,19} = 6.84$ 27 or $\chi_0^2 > \chi^2_{\alpha,2,n-1}$ where $\alpha = 0.01$ and $\chi^2_{0.005,19} = 38.58$ for n = 20
6) n = 20, s = 0.25

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{19(0.25)^2}{(0.10)^2} = 118.75$$

7) Because 118.75 > 38.58 reject H_0. The true standard deviation of Izod impact strength differs from 0.10 at $\alpha = 0.01$.

b) P-value < 0.005

c) 99% confidence interval for σ. First find the confidence interval for σ^2 :
For $\alpha = 0.01$ and n = 20, $\chi^2_{\alpha/2,n-1} = \chi^2_{0.995,19} = 6.84$ and $\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.005,19} = 38.58$

$$\frac{19(0.25)^2}{38.58} \le \sigma^2 \le \frac{19(0.25)^2}{6.84}$$
$$0.03078 \le \sigma^2 \le 0.1736$$
$$0.175 < \sigma < 0.417$$

Because 0.01 falls below the lower confidence bound, we conclude that the population standard deviation is not equal to 0.01.

9-87

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the standard deviation of tire life, σ. However, the answer can be found by performing a hypothesis test on σ^2.
2) $H_0: \sigma^2 = 4000^2$
3) $H_1: \sigma^2 < 4000^2$
4) $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$
5) Reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.95,15} = 7.26$ for n = 16
6) n = 16, $s^2 = (3645.94)^2$

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{15(3645.94)^2}{4000^2} = 12.46$$

7) Because 12.46 > 7.26 fail to reject H_0. There is not sufficient evidence to conclude the true standard deviation of tire life is less than 4000 km at $\alpha = 0.05$.

P-value = P(χ^2 <12.46) for 15 degrees of freedom. Thus, 0.5 < 1-P-value < 0.9 and 0.1 < P-value < 0.5

b) The 95% one sided confidence interval below includes the value 4000. Therefore, we are not able to conclude that the variance is less than 4000^2.

$$\sigma^2 \le \frac{15(3645.94)^2}{7.26} = 27464625$$
$$\sigma \le 5240$$

9-88

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true standard deviation of the diameter,σ. However, the solution can be found by performing a hypothesis test on σ^2.
2) H_0 : $\sigma^2 = 0.0001$
3) H_1 : $\sigma^2 > 0.0001$
4) $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$
5) Reject H_0 if $\chi_0^2 > \chi^2_{\alpha,n-1}$ where $\alpha = 0.01$ and $\chi^2_{0.01,14} = 29.14$ for n = 15
6) n = 15, $s^2 = 0.008$

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{14(0.008)^2}{0.0001} = 8.96$$

7) Because 8.96 < 29.14 fail to reject H_0. There is insufficient evidence to conclude that the true standard deviation of the diameter exceeds 0.0001 at $\alpha = 0.01$.

P-value = P(χ^2 > 8.96) for 14 degrees of freedom: 0.5 < P-value < 0.9

b) Using the chart in the Appendix, with $\lambda = \frac{0.015}{0.01} = 1.5$ and n = 15 we find β = 0.50.

c) $\lambda = \frac{\sigma}{\sigma_0} = \frac{0.0125}{0.01} = 1.25$, power = 0.8, β = 0.2, using Chart VII k) the required sample size is 50

9-89

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true variance of sugar content, σ^2. The answer can be found by performing a hypothesis test on σ^2.
2) H_0 : $\sigma^2 = 18$
3) H_1 : $\sigma^2 \neq 18$
4) $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$
5) Reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.975,9} = 2.70$ or $\chi_0^2 > \chi^2_{\alpha,2,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.025,9} = 19.02$ for n = 10
6) n = 10, s = 4.8

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{9(4.8)^2}{18} = 11.52$$

7) Because 11.52 < 19.02 fail to reject H_0. The true variance of sugar content differs from 18 at $\alpha = 0.01$.The χ_0^2 is between 0.10 and 0.50. Therefore, 0.2 < P-value < 1

b) Using the chart in the Appendix, with $\lambda = 2$ and n = 10, β = 0.45.

c) Using the chart in the Appendix, with $\lambda = \sqrt{\frac{40}{18}} = 1.49$ and β = 0.10, n = 30.

Section 9-5

9-90 a) A two-sided test because the alternative hypothesis is p not = 0.4

b) sample p = $\dfrac{X}{N} = \dfrac{98}{275} = 0.3564$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{98 - 275(0.4)}{\sqrt{275(0.4)(0.6)}} = -1.4771$$

P-value = 2(1 – Φ(1.4771)) = 2(1 – 0.9302) = 0.1396

c) The normal approximation is appropriate because np > 5 and n(1 - p) > 5.

9-91 a) A one-sided test because the alternative hypothesis is p < 0.6

b) The test is based on the normal approximation. It is appropriate because np > 5and n(1 - p) > 5.

c) sample p = $\dfrac{X}{N} = \dfrac{287}{500} = 0.574$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{287 - 500(0.6)}{\sqrt{500(0.6)(0.4)}} = -1.1867$$

P-value = Φ(-1.1867) = 0.1177

The 95% upper confidence interval is:

$$p \le \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$p \le 0.574 + 1.65\sqrt{\frac{0.574(0.426)}{500}}$$

$$p \le 0.6105$$

d) P-value = 2[1 - Φ(1.1867)] = 2(1 - 0.8823) = 0.2354

9-92 a)

1) The parameter of interest is the true fraction of satisfied customers.

2) $H_0 : p = 0.9$

3) $H_1 : p \neq 0.9$

4) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$;

Either approach will yield the same conclusion

5) Reject H_0 if $z_0 < -z_{\alpha/2}$ where $\alpha = 0.05$ and $-z_{\alpha/2} = -z_{0.025} = -1.96$ or $z_0 > z_{\alpha/2}$ where $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$

6) x = 850 n = 1000 $\hat{p} = \dfrac{850}{1000} = 0.85$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{850 - 1000(0.9)}{\sqrt{1000(0.9)(0.1)}} = -5.27$$

7) Because -5.27 < -1.96 reject the null hypothesis and conclude the true fraction of satisfied customers differs from 0.9 at α = 0.05.

The P-value: 2(1-Φ(5.27)) ≤ 2(1-1) ≈ 0

b) The 95% confidence interval for the fraction of surveyed customers is:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.85 - 1.96\sqrt{\frac{0.85(0.15)}{1000}} \le p \le .85 + 1.96\sqrt{\frac{0.85(0.15)}{1000}}$$

$$0.827 \le p \le 0.87$$

Because 0.9 is not included in the confidence interval, reject the null hypothesis at $\alpha = 0.05$.

9-93 a)

1) The parameter of interest is the true fraction of rejected parts

2) $H_0 : p = 0.03$

3) $H_1 : p < 0.03$

4) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion

5) Reject H_0 if $z_0 < -z_\alpha$ where $\alpha = 0.05$ and $-z_\alpha = -z_{0.05} = -1.65$

6) $x = 10$ $n = 500$ $\hat{p} = \dfrac{10}{500} = 0.02$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{10 - 500(0.03)}{\sqrt{500(0.03)(0.97)}} = -1.31$$

7) Because $-1.31 > -1.65$ fail to reject the null hypothesis. There is not enough evidence to conclude that the true fraction of rejected parts is less than 0.03 at $\alpha = 0.05$. P-value = $\Phi(-1.31) = 0.095$

b)

The upper one-sided 95% confidence interval for the fraction of rejected parts is:

$$p \le \hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$p \le .02 + 1.65\sqrt{\frac{0.02(0.98)}{500}}$$

$$p \le 0.0303$$

Because $0.03 < 0.0303$, we fail to reject the null hypothesis

9-94 a)

1) The parameter of interest is the true fraction defective integrated circuits

2) $H_0 : p = 0.05$

3) $H_1 : p \ne 0.05$

4) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion

5) Reject H_0 if $z_0 < -z_{\alpha/2}$ where $\alpha = 0.05$ and $-z_{\alpha/2} = -z_{0.025} = -1.96$ or $z_0 > z_{\alpha/2}$ where $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$

6) $x = 13$ $n = 300$ $\hat{p} = \dfrac{13}{300} = 0.043$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{13 - 300(0.05)}{\sqrt{300(0.05)(0.95)}} = -0.53$$

7) Because −0.53 > −1.65 fail to reject null hypothesis. The fraction of defective integrated circuits is not significantly different from 0.05, at $\alpha = 0.05$.

P-value = 2(1-Φ(0.53)) = 2(1-0.70194) = 0.59612

b) The 95% confidence interval is:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.043 - 1.96\sqrt{\frac{0.043(0.957)}{300}} \le p \le 043 + 1.96\sqrt{\frac{0.043(0.957)}{300}}$$

$$0.02004 \le p \le 0.065$$

Because the hypothesized value (p = 0.05) is contained in the confidence interval we fail to reject the null hypothesis.

9-95 a)

1) The parameter of interest is the true success rate

2) $H_0 : p = 0.78$

3) $H_1 : p > 0.78$

4) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion

5) Reject H_0 if $z_0 > z_\alpha$. Since the value for α is not given. We assume $\alpha = 0.05$ and $z_\alpha = z_{0.05} = 1.65$

6) $x = 289$ $n = 350$ $\hat{p} = \dfrac{289}{350} \cong 0.83$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{289 - 350(0.78)}{\sqrt{350(0.78)(0.22)}} = 2.06$$

7) Because 2.06 > 1.65 reject the null hypothesis and conclude the true success rate is greater than 0.78, at $\alpha = 0.05$.

P-value = 1 - 0.9803 = 0.0197

b) The 95% lower confidence interval:

$$\hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

$$.83 - 1.65\sqrt{\frac{0.83(0.17)}{350}} \le p$$

$$0.7969 \le p$$

Because the hypothesized value is not in the confidence interval (0.78 < 0.7969), reject the null hypothesis.

9-96 a)

1) The parameter of interest is the true percentage of polished lenses that contain surface defects, p.

2) $H_0 : p = 0.02$

3) $H_1 : p < 0.02$

4) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion

5) Reject H_0 if $z_0 < -z_\alpha$ where $\alpha = 0.05$ and $-z_\alpha = -z_{0.05} = -1.65$

6) x = 6 n = 250 $\hat{p} = \dfrac{6}{250} = 0.024$

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.024 - 0.02}{\sqrt{\frac{0.02(1-0.02)}{250}}} = 0.452$$

7) Because 0.452 > –1.65 fail to reject the null hypothesis. There is not sufficient evidence to qualify the machine at the 0.05 level of significance. P-value = $\Phi(0.452) = 0.67364$

b) The upper 95% confidence interval is:

$$p \le \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$p \le .024 + 1.65\sqrt{\frac{0.024(0.976)}{250}}$$
$$p \le 0.0264$$

Because the confidence interval contains the hypothesized value ($p = 0.02 \le 0.0264$) we fail to reject the null hypothesis.

9-97 a)

1) The parameter of interest is the true percentage of football helmets that contain flaws, p.
2) $H_0 : p = 0.1$
3) $H_1 : p > 0.1$
4) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion

5) Reject H_0 if $z_0 > z_\alpha$ where $\alpha = 0.01$ and $z_\alpha = z_{0.01} = 2.33$

6) x = 16 n = 200 $\hat{p} = \dfrac{16}{200} = 0.08$

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.08 - 0.10}{\sqrt{\frac{0.10(1-0.10)}{200}}} = -0.94$$

7) Because –0.94 < 2.33 fail to reject the null hypothesis. There is not enough evidence to conclude that the proportion of football helmets with flaws exceeds 10%.
P-value = $1 - \Phi(-0.94) = 0.8264$

b) The 99% lower confidence interval

$$\hat{p} - z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

$$.08 - 2.33\sqrt{\frac{0.08(0.92)}{200}} \le p$$

$$0.035 \le p$$

Because the confidence interval contains the hypothesized value ($0.035 \le p = 0.1$) we fail to reject the null hypothesis.

9-98 a)

1) The parameter of interest is the true proportion of engineering students planning graduate studies
2) $H_0 : p = 0.50$
3) $H_1 : p \ne 0.50$
4) $z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion
5) Reject H_0 if $z_0 < -z_{\alpha/2}$ where $\alpha = 0.05$ and $-z_{\alpha/2} = -z_{0.025} = -1.96$ or $z_0 > z_{\alpha/2}$ where $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$
6) x = 117 n = 484 $\hat{p} = \frac{117}{484} = 0.2423$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{117 - 484(0.5)}{\sqrt{484(0.5)(0.5)}} = -11.36$$

7) Because $-11.36 > -1.65$ reject the null hypothesis and conclude that the true proportion of engineering students planning graduate studies differs from 0.5, at $\alpha = 0.05$.

P-value = $2[1 - \Phi(11.36)] \cong 0$

b) $\hat{p} = \frac{117}{484} = 0.242$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.242 - 1.96\sqrt{\frac{0.242(0.758)}{484}} \le p \le 0.242 - 1.96\sqrt{\frac{0.242(0.758)}{484}}$$

$$0.204 \le p \le 0.280$$

Because the 95% confidence interval does not contain the value 0.5 we conclude that the true proportion of engineering students planning graduate studies differs from 0.5.

9-99 1) The parameter of interest is the true proportion of batteries that fail before 48 hours, p.
2) $H_0 : p = 0.002$
3) $H_1 : p < 0.002$
4) $z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion
5) Reject H_0 if $z_0 < -z_{\alpha}$ where $\alpha = 0.01$ and $-z_{\alpha} = -z_{0.01} = -2.33$

6) x = 15 n = 5000 $\hat{p} = \dfrac{15}{5000} = 0.003$

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.003 - 0.002}{\sqrt{\dfrac{0.002(1-0.998)}{5000}}} = 1.58$$

7) Because 1.58 > -2.33 fail to reject the null hypothesis. There is not sufficient evidence to conclude that the proportion of cell phone batteries that fail is less than 0.2% at $\alpha = 0.01$.

9-100. The problem statement implies that H_0: p = 0.6, H_1: p > 0.6 and defines an acceptance region as $\hat{p} \le \dfrac{315}{500} = 0.63$ and rejection region as $\hat{p} > 0.63$

a) The probability of a type 1 error is

$$\alpha = P(\hat{p} \ge 0.63 \mid p = 0.6) = P\left(Z \ge \frac{0.63 - 0.6}{\sqrt{\dfrac{0.6(0.4)}{500}}} \right) = P(Z \ge 1.37) = 1 - P(Z < 1.37) = 0.08535.$$

b) $\beta = P(\hat{P} \le 0.63 \mid p = 0.75) = P(Z \le -6.196) = 0$.

9-101 a)

1) The parameter of interest is the true proportion of engine crankshaft bearings exhibiting surface roughness.

2) $H_0 : p = 0.10$

3) $H_1 : p > 0.10$

4) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ or $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$; Either approach will yield the same conclusion

5) Reject H_0 if $z_0 > z_\alpha$ where $\alpha = 0.05$ and $z_\alpha = z_{0.05} = 1.65$

6) $x = 10$ $n = 85$ $\hat{p} = \dfrac{10}{85} = 0.118$

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{10 - 85(0.10)}{\sqrt{85(0.10)(0.90)}} = 0.54$$

7) Because 0.54 < 1.65 fail to reject the null hypothesis. There is not enough evidence to conclude that the true proportion of crankshaft bearings exhibiting surface roughness exceeds 0.10, at $\alpha = 0.05$.

P-value = 1 – $\Phi(0.54)$ = 0.295

b) $p = 0.15$, $p_0 = 0.10$, $n = 85$, and $z_{\alpha/2}=1.96$

$$\beta = \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}} \right) - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}} \right)$$

$$= \Phi\left(\frac{0.10 - 0.15 + 1.96\sqrt{0.10(1-0.10)/85}}{\sqrt{0.15(1-0.15)/85}} \right) - \Phi\left(\frac{0.10 - 0.15 - 1.96\sqrt{0.10(1-0.10)/85}}{\sqrt{0.15(1-0.15)/85}} \right)$$

$$= \Phi(0.36) - \Phi(-2.94) = 0.6406 - 0.0016 = 0.639$$

c) $n = \left(\dfrac{z_{\alpha/2}\sqrt{p_0(1-p_0)} - z_\beta\sqrt{p(1-p)}}{p-p_0} \right)^2$

$$n = \left(\frac{1.96\sqrt{0.10(1-0.10)} - 1.28\sqrt{0.15(1-0.15)}}{0.15-0.10} \right)^2$$

$$= (10.85)^2 = 117.63 \cong 118$$

9-102 1) The parameter of interest is the true percentage of customers received fully charged batteries, p.
2) $H_0 : p = 0.85$
3) $H_1 : p > 0.85$
4)

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \quad \text{or} \quad z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Either approach will yield the same conclusion

5) Reject H_0 if $z_0 > z_\alpha$ where $\alpha = 0.05$ and $z_{0.05} = 1.65$

6) $x = 96$ $n = 100$

$$\hat{p} = \frac{96}{100} = 0.96$$

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.96 - 0.85}{\sqrt{\frac{0.85(1-0.85)}{100}}} = 3.08$$

7) Because 3.08 > 1.65 reject the null hypothesis. There is enough evidence to conclude that the proportion of customers received fully charged batteries is at least as high as the previous model.

9-103

1) The parameter of interest is the rate of zip codes that can be correctly read, p.
2) $H_0 : p = 0.90$
3) $H_1 : p > 0.90$
4)

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \quad \text{or} \quad z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Either approach will yield the same conclusion

5) Reject H_0 if $z_0 > z_\alpha$ where $\alpha = 0.05$ and $z_{0.05} = 1.65$

6) $x = 466$ $n = 500$

$$\hat{p} = \frac{466}{500} = 0.932$$

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.932 - 0.90}{\sqrt{\frac{0.90(1-0.90)}{500}}} = 2.385$$

7) Because 2.385 > 1.65 reject the null hypothesis. There is enough evidence to conclude that the rate of zip codes that can be correctly read is at least 90% at $\alpha = 0.05$.

9-104

The 90% lower confidence interval

$$\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

$$0.932 - 1.28\sqrt{\frac{0.932(0.068)}{500}} \le p$$

$$0.9176 \le p$$

The 90% confidence interval is $0.9176 < p$. Because 0.90 does not lie in this interval, we reject the null hypothesis. As a result, the confidence interval supports the claim that at least 90% of the zip codes can be correctly read.

9-105

The 95% lower confidence interval

$$\hat{p} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p$$

$$0.826 - 1.65\sqrt{\frac{0.826(0.174)}{350}} \le p$$

$$0.792 \le p$$

The 95% confidence interval is $0.792 < p$. Because 0.78 does not lie in this interval, the confidence interval supports the claim that at least 78% of the procedures are successful.

Section 9-7

9-106 The expected frequency is found from the Poisson distribution $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ with λ = 1.2

Value	0	1	2	3	4 or more
Observed Frequency	24	30	31	11	4
Expected Frequency	30.12	36.14	21.69	8.67	3.37

Because the expected frequencies of all categories are greater than 3, there is no need to combine categories.

The degrees of freedom are $k - p - 1 = 5 - 0 - 1 = 4$

a)
1) Interest is on the form of the distribution for X.
2) H_0: The form of the distribution is Poisson
3) H_1: The form of the distribution is not Poisson
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_o^2 > \chi_{0.05,4}^2 = 9.49$ for α = 0.05
6)

$$\chi_0^2 = \frac{(24-30.12)^2}{30.12} + \frac{(30-36.14)^2}{36.14} + \frac{(31-21.69)^2}{21.69} + \frac{(11-8.67)^2}{8.67} + \frac{(4-3.37)^2}{3.37} = 7.0267$$

7) Because 7.0267 < 9.49 fail to reject H_0.

b) The P-value is between 0.1 and 0.2 using Table IV. From computer software, the P-value = 0.1345

9-107

The expected frequency is found from the Poisson distribution $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ where

$\lambda = [1(1) + 2(11) + \ldots + 7(10) + 8(9)]/75 = 4.907$ is the estimated mean.

Value	1 or less	2	3	4	5	6	7	8 or more
Observed Frequency	1	11	8	13	11	12	10	9
Expected Frequency	3.2760	6.6770	10.9213	13.3977	13.1485	10.7533	7.5381	9.2880

Because the expected frequencies of all categories are greater than 3, there is no need to combine categories.

The degrees of freedom are k – p – 1 = 8 – 1 – 1 = 6

a)
1) Interest is on the form of the distribution for the number of flaws.
2) H_0: The form of the distribution is Poisson
3) H_1: The form of the distribution is not Poisson
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_o^2 > \chi_{0.01,6}^2 = 16.81$ for α = 0.01
6)

$$\chi_0^2 = \frac{(1-3.2760)^2}{3.2760} + \frac{(11-6.6770)^2}{6.6770} + \cdots + \frac{(9-9.2880)^2}{9.2880} = 6.482$$

7) Because 6.482 < 16.81 fail to reject H_0.

b) P-value = 0.3715 (from computer software)

9-108 Estimated mean = 10.131

Value	**5 or less**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15 or more**
Rel. Freq	0.067	0.067	0	0.100	0.133	0.200	0.133	0.133	0.067	0.033	0.067
Observed (Days)	2	2	0	3	4	6	4	4	2	1	2
Expected (Days)	1.8687	1.7942	2.5967	3.2884	3.7016	3.7501	3.4538	2.9159	2.2724	1.6444	2.7138

Because there are several cells with expected frequencies less than 3, a revised table follows. We note that there are other reasonable alternatives to combine cells. For example, cell 7 could also be combined with cell 8.

Value	**7 or less**	**8**	**9**	**10**	**11**	**12 or more**
Observed (Days)	4	3	4	6	4	9
Expected (Days)	6.2596	3.2884	3.7016	3.7501	3.4538	9.5465

The degrees of freedom are k – p – 1 = 6 – 1 – 1 = 4

a)
1) Interest is on the form of the distribution for the number of calls arriving to a switchboard from noon to 1pm during business days.
2) H_0: The form of the distribution is Poisson
3) H_1: The form of the distribution is not Poisson
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_o^2 > \chi^2_{0.05,4} = 9.49$ for $\alpha = 0.05$
6)

$$\chi_0^2 = \frac{(4-6.2596)^2}{6.2596} + \ldots + \frac{(9-9.5465)^2}{9.5465} = 2.2779$$

7) Because 2.2779 < 9.49, fail to reject H_0.

b) The P-value is between 0.5 and 0.9 using Table IV. P-value = 0.685 (from computer software)

9-109

Under the null hypothesis there are 50 observations from a binomial distribution with n = 6 and p = 0.25. Use the binomial distribution to obtain the expected frequencies from the 50 observations.

Value	0	1	2	3	4 or more
Observed	4	21	10	13	2
Expected	8.8989	17.7979	14.8315	6.5918	1.8799

The expected frequency for cell "4 or more" is less than 3. Combine this cell with its neighboring cell to obtain the following table.

Value	0	1	2	3 or more

Observed	4	21	10	15
Expected	8.8989	17.7979	14.8315	8.4717

The degrees of freedom are k – p – 1 = 4 – 0 – 1 = 3

a)
1) The variable of interest is the form of the distribution for the random variable X.
2) H_0: The form of the distribution is binomial with n = 6 and p = 0.25
3) H_1: The form of the distribution is not binomial with n = 6 and p = 0.25
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_o^2 > \chi_{0.05,3}^2 = 7.81$ for $\alpha = 0.05$

6)
$$\chi_0^2 = \frac{(4-8.8989)^2}{8.8989} + \cdots + \frac{(15-8.2397)^2}{8.2397} = 9.8776$$

7) Because 9.8776 > 7.81 reject H_0. We conclude that the distribution is not binomial with n = 6 and p = 0.25 at $\alpha = 0.05$.

b) P-value = 0.0197 (from computer software)

9-110

Under the null hypothesis there are 75 observations from a binomial distribution with n = 24. The value of *p* must be estimated. Let the estimate be denoted by $\hat{p}_{\text{sample}}$. The sample mean $= \frac{0(39)+1(23)+2(12)+3(1)}{75} = 0.6667$ and the mean of a binomial distribution is np. Therefore,
$\hat{p}_{sample} = \frac{sample\ \ mean}{n} = \frac{0.6667}{24} = 0.02778$. From the binomial distribution with n = 24 and p = 0.02778 the expected frequencies follow

Value	0	1	2	3 or more
Observed	39	23	12	1
Expected	38.1426	26.1571	8.5952	2.1051

Because the cell “3 or more” has an expected frequency less than 3, combine this category with that of the neighboring cell to obtain the following table.

Value	0	1	2 or more
Observed	39	23	13
Expected	38.1426	26.1571	10.7003

The degrees of freedom are k – p – 1 = 3 – 1 – 1 = 1

a)
1) Interest is the form of the distribution for the number of underfilled cartons, X.
2) H_0: The form of the distribution is binomial
3) H_1: The form of the distribution is not binomial
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_o^2 > \chi_{0.05,1}^2 = 3.84$ for $\alpha = 0.05$

6) $$\chi_0^2 = \frac{(39-38.1426)^2}{38.1426} + \frac{(23-26.1571)^2}{26.1571} + \frac{(13-10.7003)^2}{10.7003} = 0.8946$$

7) Because 0.8946 < 3.84 fail to reject H_0.

b) The P-value is between 0.3 and 0.5 using Table IV. From Minitab the P-value = 0.3443.

9-111

The estimated mean = 49.6741. Based on a Poisson distribution with λ = 49.674 the expected frequencies are shown in the following table. All expected frequencies are greater than 3.

The degrees of freedom are k – p – 1 = 26 – 1 – 1 = 24

Vechicles per minute	Frequency	Expected Frequency
40 or less	14	277.6847033
41	24	82.66977895
42	57	97.77492539
43	111	112.9507307
44	194	127.5164976
45	256	140.7614945
46	296	152.0043599
47	378	160.6527611
48	250	166.2558608
49	185	168.5430665
50	171	167.4445028
51	150	163.091274
52	110	155.7963895
53	102	146.0197251
54	96	134.3221931
55	90	121.3151646
56	81	107.6111003
57	73	93.78043085
58	64	80.31825
59	61	67.62265733
60	59	55.98491071
61	50	45.5901648
62	42	36.52661944
63	29	28.80042773
64	18	22.35367698
65 or more	15	62.60833394

a)
1) Interest is the form of the distribution for the number of cars passing through the intersection.
2) H_0: The form of the distribution is Poisson
3) H_1: The form of the distribution is not Poisson
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_o^2 > \chi_{0.05,24}^2 = 36.42$ for $\alpha = 0.05$
6) Estimated mean = 49.6741

$$\chi_0^2 = 1012.8044$$

7) Because 1012.804351 >> 36.42, reject H_0. We can conclude that the distribution is not a Poisson distribution at α = 0.05.

b) P-value ≈ 0 (from computer software)

9-112

The expected frequency is determined by using the Poisson distribution

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$ where $\lambda = [6(1) + 7(1) + \ldots + 39(1) + 41(1)]/110 = 19.2455$ is the estimated mean.

The expected frequencies are shown in the following table.

Number of Earthquakes	Frequency	Expected Frequency
6 or less	1	0.048065
7	1	0.093554
8	4	0.225062
9	0	0.4812708
10	3	0.926225
11	4	1.620512
12	3	2.598957
13	6	3.847547
14	5	5.289128
15	11	6.786111
16	8	8.162612
17	3	9.240775
18	9	9.880162
19	4	10.0078
20	4	9.630233
21	7	8.82563
22	8	7.720602
23	4	6.460283
24	3	5.180462
25	2	3.988013
26	4	2.951967
27	4	2.104146
28	1	1.446259
29	1	0.95979
30	1	0.61572
31	1	0.382252
32	2	0.229894
33	0	0.134073
34	1	0.075891
35	1	0.04173
36	2	0.022309
37	0	0.011604
38	0	0.005877
39	1	0.0029

40	0	.001395
41 or more	1	0.001191

After combining categories with frequencies less than 3, we obtain the following table. We note that there are other reasonable alternatives to combine cells. For example, cell 12 could also be combined with cell 13.

Number of earthquakes	Frequency	Expected Frequency	Chi squared
12 or less	16	5.993646	16.70555
13	6	3.847547	1.204158
14	5	5.289128	0.015805
15	11	6.786111	2.616648
16	8	8.162612	0.003239
17	3	9.240775	4.214719
18	9	9.880162	0.078408
19	4	10.0078	3.606553
20	4	9.630233	3.291667
21	7	8.82563	0.377642
22	8	7.720602	0.010111
23	4	6.460283	0.936955
24	3	5.180462	0.917759
25	2	3.988013	0.991019
26 or more	20	8.986998	13.49574

The degrees of freedom are k – p – 1 = 15 – 1 – 1 = 13

a)
1) Interest is the form of the distribution for the number of earthquakes per year of magnitude 7.0 and greater.
2) H_0: The form of the distribution is Poisson
3) H_1: The form of the distribution is not Poisson
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_o^2 > \chi_{0.05,13}^2 = 22.36$ for $\alpha = 0.05$
6)

$$\chi_0^2 = \frac{(16-5.9936)^2}{5.9936} + \cdots + \frac{(20-8.9856)^2}{8.9856} = 48.4660$$

7) Because 48.4660 > 22.36 reject H_0. We conclude that the distribution of the number of earthquakes is not a Poisson distribution.

b) P-value ≈ 0 (from computer software)

Section 9-8

9-113

1) Interest is on the species distribution.
2) H_0: Species distribution is independent of year.

3) H_1: Species distribution is not independent of year.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

5) The critical value is $\chi^2_{0.05,4} = 9.488$ for α = 0.05.

6) The calculated test statistic is $\chi_0^2 = 146.3648$

7) Because $\chi_0^2 > \chi^2_{0.05,4}$, reject H_0. The evidence is sufficient to claim that species distribution is not independent of year at α = 0.05. P-value = $P(\chi_0^2 > 146.3648) \approx 0$

9-114

1) Interest is on the survival distrbution.
2) H_0: Survival is independent of ticket class.
3) H_1: Survival is not independent of ticket class.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

5) The critical value is $\chi^2_{0.05,3} = 7.815$ for α = 0.05.

6) The calculated test statistic is $\chi_0^2 = 187.7932$

7) Because $\chi_0^2 > \chi^2_{0.05,3}$, reject H_0. The evidence is sufficient to claim that survival is not independent of ticket class at α = 0.05. P-value = $P(\chi_0^2 > 187.7932) \approx 0$

9-115 1) Interest is on the distribution of breakdowns among shift.
2) H_0: Breakdowns are independent of shift.
3) H_1: Breakdowns are not independent of shift.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

5) The critical value is $\chi^2_{.05,6} = 12.592$ for α = 0.05

6) The calculated test statistic is $\chi_0^2 = 11.65$

7) Because $\chi_0^2 \not> \chi^2_{0.05,6}$ fail to reject H_0. The evidence is not sufficient to claim that machine breakdown and shift are dependent at α = 0.05. P-value = $P(\chi_0^2 > 11.65)$ = 0.070 (from computer software)

9-116 1) Interest is on the distribution of calls by surgical-medical patients.
2) H_0: Calls by surgical-medical patients are independent of Medicare status.
3) H_1: Calls by surgical-medical patients are not independent of Medicare status.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

5) The critical value is $\chi^2_{.01,1} = 6.637$ for α = 0.01

6) The calculated test statistic is $\chi^2_0 = 0.033$

7) Because $\chi_0^2 \not> \chi^2_{0.01,1}$ fail to reject H_0. The evidence is not sufficient to claim that surgical-medical patients and Medicare status are dependent. *P*-value = $P(\chi_0^2 > 0.033)$ = 0.85

9-117 1) Interest is on the distribution of statistics and OR grades.
2) H_0: Statistics grades are independent of OR grades.
3) H_1: Statistics and OR grades are not independent.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

5) The critical value is $\chi^2_{.01,9} = 21.665$ for α = 0.01

6) The calculated test statistic is $\chi^2_0 = 25.55$

7) $\chi_0^2 > \chi^2_{0.01,9}$ Therefore, reject H_0 and conclude that the grades are not independent at α = 0.01.

P-value = $P(\chi_0^2 > 25.55)$ = 0.002 (from computer software)

9-118 1) Interest is on the distribution of deflections.
2) H_0: Deflection and range are independent.
3) H_1: Deflection and range are not independent.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

5) The critical value is $\chi^2_{0.05,4} = 9.488$ for α = 0.05

6) The calculated test statistic is $\chi^2_0 = 2.46$

7) Because $\chi_0^2 \not> \chi^2_{0.05,4}$ fail to reject H_0. The evidence is not sufficient to claim that the data are dependent at α = 0.05. The *P*-value = $P(\chi_0^2 > 2.46)$ = 0.652 (from computer software).

9-119 1) Interest is on the distribution of failures of an electronic component.
2) H_0: Type of failure is independent of mounting position.
3) H_1: Type of failure is not independent of mounting position.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

5) The critical value is $\chi^2_{0.01,3} = 11.344$ for α = 0.01

6) The calculated test statistic is $\chi^2_0 = 10.71$

7) Because $\chi_0^2 \not> \chi_{0.01,3}^2$ fail to reject H_0. The evidence is not sufficient to claim that the type of failure is dependent on the mounting position at $\alpha = 0.01$. *P*-value = $P(\chi_0^2 > 10.71) = 0.013$ (from computer software).

9-120 1) Interest is on the distribution of opinion on core curriculum change.
2) H_0: Opinion of the change is independent of the class standing.
3) H_1: Opinion of the change is not independent of the class standing.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

5) The critical value is $\chi^2_{0.05,3} = 7.815$ for $\alpha = 0.05$

6) The calculated test statistic is $\chi^2_{.0} = 26.97$.

7) $\chi_0^2 >>> \chi^2_{0.05,3}$, reject H_0 and conclude that opinion on the change and class standing are not independent. *P*-value = $P(\chi_0^2 > 26.97) \approx 0$

9-121 a)
1) Interest if on the distribution of successes.
2) H_0: successes are independent of size of stone.
3) H_1: successes are not independent of size of stone.
4) The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

5) The critical value is $\chi^2_{.05,1} = 3.84$ for $\alpha = 0.05$

6) The calculated test statistic χ^2_0 = 13.766 with details below.

7) $\chi_0^2 > \chi^2_{0.05,1}$, reject H_0 and conclude that the number of successes and the stone size are not independent.

```
            1      2     All
1          55     25      80
        66.06  13.94   80.00
2         234     36     270
       222.94  47.06  270.00
All       289     61     350
       289.00  61.00  350.00
Cell Contents:      Count
                    Expected count
Pearson Chi-Square = 13.766, DF = 1, P-Value = 0.000
```

b) *P*-value = $P(\chi_0^2 > 13.766) < 0.005$

Section 9-9

9-122 a)
1) The parameter of interest is the median of pH.
2) $H_0 : \tilde{\mu} = 7.0$
3) $H_1 : \tilde{\mu} \neq 7.0$
4) The test statistic is the observed number of plus differences or $r^+ = 8$ for $\alpha = 0.05$.
5) Reject H_0 if the P-value corresponding to $r^+ = 8$ is less than or equal to $\alpha = 0.05$.
6) Using the binomial distribution with $n = 10$ and $p = 0.5$, the P-value = $2P(R^+ \geq 8 \mid p = 0.5) = 0.1$
7) We fail to reject H_0. There is not enough evidence to conclude that the median of the pH differs from7.0.

b)

1) The parameter of interest is median of pH.
2) $H_0 : \tilde{\mu} = 7.0$
3) $H_1 : \tilde{\mu} \neq 7.0$
4) The test statistic is $z_0 = \dfrac{r^* - 0.5n}{0.5\sqrt{n}}$
5) Reject *H0* if $|Z_0| > 1.96$ for α=0.05.
6) r*=8 and $z_0 = \dfrac{r^* - 0.5n}{0.5\sqrt{n}} = \dfrac{8 - 0.5(10)}{0.5\sqrt{10}} = 1.90$
7) Fail to reject *H0*. There is not enough evidence to conclude that the median of the pH differs from 7.0.

P-value = 2[1 - P($|Z_0|$ < 1.90)] = 2(0.0287) = 0.0574

9-123 a)
1) The parameter of interest is median titanium content.
2) $H_0 : \tilde{\mu} = 8.5$
3) $H_1 : \tilde{\mu} \neq 8.5$
4) The test statistic is the observed number of plus differences or $r^+ = 7$ for $\alpha = 0.05$.
5) Reject *H0* if the *P*-value corresponding to $r^+ = 7$ is less than or equal to $\alpha = 0.05$.
6) Using the binomial distribution with n = 20 and p = 0.5, *P*-value = 2P($R^* \leq 7$ | p = 0.5) = 0.1315
7) We fail to reject *H0*. There is not enough evidence to conclude that the median of the titanium content differs from 8.5.

b)
1) Parameter of interest is the median titanium content
2) $H_0 : \tilde{\mu} = 8.5$
3) $H_1 : \tilde{\mu} \neq 8.5$
4) Test statistic is $z_0 = \dfrac{r^+ - 0.5n}{0.5\sqrt{n}}$
5) Reject *H0* if the $|Z_0| > Z_{0.025} = 1.96$ for α=0.05
6) Computation: $z_0 = \dfrac{7 - 0.5(20)}{0.5\sqrt{20}} = -1.34$
7) We fail to reject *H0*. There is not enough evidence to conclude that the median titanium content differs from 8.5. The P-value = 2P(Z < -1.34) = 0.1802.

9-124 a)
1) Parameter of interest is the median impurity level.
2) $H_0 : \tilde{\mu} = 2.5$
3) $H_1 : \tilde{\mu} < 2.5$
4) The test statistic is the observed number of plus differences or $r^+ = 2$ for $\alpha = 0.05$.
5) Reject *H0* if the *P-value* corresponding to $r^+ = 2$ is less than or equal to $\alpha = 0.05$.
6) Using the binomial distribution with $n = 22$ and $p = 0.5$, the P-value = P($R^+ \leq 2$ | p = 0.5) = 0.0002
7) Conclusion, reject *H0*. The data supports the claim that the median is impurity level is less than 2.5.

b)
1) Parameter of interest is the median impurity level
2) $H_0 : \tilde{\mu} = 2.5$
3) $H_1 : \tilde{\mu} < 2.5$
4) Test statistic is $z_0 = \dfrac{r^+ - 0.5n}{0.5\sqrt{n}}$
5) Reject *H0* if the $Z_0 < Z_{0.05} = -1.65$ for α=0.05

6) Computation: $z_0 = \dfrac{2-0.5(22)}{0.5\sqrt{22}} = -3.84$

7) Reject H_0 and conclude that the median impurity level is less than 2.5.
The P-value = P(Z < -3.84) = 0.000062

9-125 a)
1) Parameter of interest is the median margarine fat content
2) $H_0 : \tilde{\mu} = 17.0$
3) $H_1 : \tilde{\mu} \neq 17.0$
4) α = 0.05
5) The test statistic is the observed number of plus differences or $r^+ = 3$.
6) Reject H_0 if the P-value corresponding to $r^+ = 3$ is less than or equal to α=0.05.
7) Using the binomial distribution with n = 6 and p = 0.5, the P-value = 2*P($R^+ \geq 3$|p=0.5,n=6) ≈ 1.
8) Fail to reject H_0. There is not enough evidence to conclude that the median fat content differs from 17.0.

b)
1) Parameter of interest is the median margarine fat content
2) $H_0 : \tilde{\mu} = 17.0$
3) $H_1 : \tilde{\mu} \neq 17.0$
4) Test statistic is $z_0 = \dfrac{r^+ - 0.5n}{0.5\sqrt{n}}$
5) Reject H_0 if the $|Z_0| > Z_{0.025} = 1.96$ for α=0.05
6) Computation: $z_0 = \dfrac{3-0.5(6)}{0.5\sqrt{6}} = 0$
7) Fail to reject H_0. The P-value = 2[1 – Φ(0)] = 2(1 – 0.5) = 1. There is not enough evidence to conclude that the median fat content differs from 17.0.

9-126 a)
1) Parameter of interest is the median compressive strength
2) $H_0 : \tilde{\mu} = 2250$
3) $H_1 : \tilde{\mu} > 2250$
4) The test statistic is the observed number of plus differences or $r^+ = 7$ for α = 0.05
5) Reject H_0 if the P-value corresponding to $r^+ = 7$ is less than or equal to α=0.05.
6) Using the binomial distribution with n = 12 and p = 0.5, the P-value = P($R^+ \geq 7$ | p = 0.5) = 0.3872
7) Fail to reject H_0. There is not enough evidence to conclude that the median compressive strength is greater than 2250.

b)
1) Parameter of interest is the median compressive strength
2) $H_0 : \tilde{\mu} = 2250$
3) $H_1 : \tilde{\mu} > 2250$
4) Test statistic is $z_0 = \dfrac{r^+ - 0.5n}{0.5\sqrt{n}}$
5) Reject H_0 if the $|Z_0| > Z_{0.025} = 1.96$ for α=0.05
6) Computation: $z_0 = \dfrac{7-0.5(12)}{0.5\sqrt{12}} = 0.577$
7) Fail to reject H_0. The P-value = 1 – Φ(0.58) = 1 – 0.7190 = 0.281. There is not enough evidence to conclude that the median compressive strength is greater than 2250.

9-127 a)
1) The parameter of interest is the mean ball diameter
2) H_0: $\mu_0 = 0.265$

3) H0: $\mu_0 \neq 0.265$

4) w = min(w$^+$, w$^-$)

5) Reject H0 if $w \leq w^*_{0.05,n=9} = 5$ for α = 0.05

6) Usually zeros are dropped from the ranking and the sample size is reduced. The sum of the positive ranks is w$^+$ = (1+4.5+4.5+4.5+4.5+8.5+8.5) = 36. The sum of the negative ranks is w$^-$ = (4.5+4.5) = 9. Therefore, w = min(36, 9) = 9.

observation	Difference xi - 0.265	Signed Rank
1	0	-
6	0	-
9	0	-
3	0.001	1
2	-0.002	-4.5
4	0.002	4.5
5	0.002	4.5
7	0.002	4.5
8	0.002	4.5
12	-0.002	-4.5
10	0.003	8.5
11	0.003	8.5

7) Conclusion: because w$^-$ = 9 is not less than or equal to the critical value $w^*_{0.05,n=9} = 5$, we fail to reject the null hypothesis that the mean ball diameter is 0.265 at the 0.05 level of significance.

b)

$$Z_0 = \frac{W^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} = \frac{36 - 9(10)/4}{\sqrt{9(10)(19)/24}} = 1.5993$$

and $Z_{0.025}$ = 1.96. Because Z_0 = 1.5993< $Z_{0.025}$ = 1.96 we fail to reject the null hypothesis that the mean ball diameter is 0.265 at the 0.05 level of significance. Also, the P-value = 2[1 – P(Z_0 < 1.5993)] = 0.1098.

9-128 1) The parameter of interest is mean hardness

2) H0: $\mu_0 = 60$

3) H0: $\mu_0 > 60$

4) w$^-$

5) Reject H0 if $w^- \leq w^*_{0.05,n=7} = 3$ for α = 0.05

6) The sum of the positive rank is w$^+$ = (3.5+5.5) = 9. The sum of the negative rank is w$^-$ = (1+2+3.5+5.5+7) = 19.

observation	Difference xi - 60	Sign Rank
4	0	-
8	-1	-1
3	-2	-2
1	3	3.5
6	-3	-3.5
2	5	5.5
5	-5	-5.5
7	-7	-7

7) Conclusion: Because w$^-$ = 19 is not less than or equal to the critical value $w^*_{0.05,n=7} = 3$, we fail to reject the null hypothesis that the mean hardness reading is greater than 60.

9-129 1) The parameter of interest is the mean dying time of the primer

2) H0: $\mu_0 = 1.5$

3) H0: $\mu_0 > 1.5$

4) w^-

5) Reject H0 if $w^- \le w^*_{0.05,n=17} = 41$ for $\alpha = 0.05$

6) The sum of the positive rank is w^+ = (4+4+4+4+4+9.5+9.5+13.5+13.5+13.5+13.5+16.5+16.5) = 126. The sum of the negative rank is w^- = (4+4+9.5+9.5) = 27.

Observation	Difference xi - 1.5	Sign Rank
1.5	0	-
1.5	0	-
1.5	0	-
1.6	0.1	4
1.6	0.1	4
1.6	0.1	4
1.4	-0.1	-4
1.6	0.1	4
1.4	-0.1	-4
1.6	0.1	4
1.3	-0.2	-9.5
1.7	0.2	9.5
1.7	0.2	9.5
1.3	-0.2	-9.5
1.8	0.3	13.5
1.8	0.3	13.5
1.8	0.3	13.5
1.8	0.3	13.5
1.9	0.4	16.5
1.9	0.4	16.5

7) Conclusion: Because w^- = 27 is less than the critical value $w^*_{0.05,n=17} = 41$, we reject the null hypothesis that the mean dying time of the primer exceeds 1.5.

Section 9-10

9-130

a)

1) The parameter of interest is the mean absorption rate of the new product, μ.

2, 3) The null and alternative hypotheses that must be tested are as follows (δ = 0.50):

$H_0 : \mu = 18.50$ and $H_0 : \mu = 17.50$

$H_1 : \mu \le 18.50$ $H_1 : \mu \ge 17.50$

b)

Test the first hypothesis ($H_0 : \mu = 18.50$ vs. $H_1 : \mu \le 18.50$):

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

5) Reject H_0 if $t_0 < -t_{\alpha,n-1}$ where $t_{0.05,19} = 1.729$ for n = 20 and $\alpha = 0.05$.

6) $\bar{x} = 18.22$, $s = 0.92$, n = 20

$$t_0 = \frac{18.22 - 18.50}{0.92/\sqrt{20}} = -1.361$$

7) Because -1.361 > -1.729, fail to reject H_0.

Test the second hypothesis ($H_0 : \mu = 17.50$ vs. $H_1 : \mu \geq 17.50$):

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,19} = 1.729$ for n = 20 and $\alpha = 0.05$.

6) $\bar{x} = 18.22$, $s = 0.92$, n = 20

$$t_0 = \frac{18.22 - 17.50}{0.92/\sqrt{20}} = 3.450$$

7) Because 3.450 > 1.729, reject H_0.

There is enough evidence to conclude that the mean absorption rate is greater than 17.50. However, there is not enough evidence to conclude that it is less than 18.50. As a result, we cannot conclude that the new product has an absorption rate that is equivalent to the absorption rate of the current one at $\alpha = 0.05$.

9-131

a)

1) The parameter of interest is the mean molecular weight of a raw material from a new supplier, μ.

2, 3) The null and alternative hypotheses that must be tested are as follows (δ=50):

$H_0 : \mu = 3550$ and $H_0 : \mu = 3450$

$H_1 : \mu \leq 3550$ $H_1 : \mu \geq 3450$

b)

Test the first hypothesis ($H_0 : \mu = 3550$ vs. $H_1 : \mu \leq 3550$):

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

5) Reject H_0 if $t_0 < -t_{\alpha,n-1}$ where $t_{0.05,9} = 1.833$ for n = 10 and $\alpha = 0.05$.

6) $\bar{x} = 3550$, $s = 25$, n = 10

$$t_0 = \frac{3550 - 3550}{25/\sqrt{10}} = 0$$

7) Because 0 > -1.833, fail to reject H_0.

Test the second hypothesis ($H_0 : \mu = 3450$ vs. $H_1 : \mu \geq 3450$):

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,9} = 1.833$ for n = 10 and $\alpha = 0.05$.

6) $\bar{x} = 3550$, $s = 25$, n = 10

$$t_0 = \frac{3550 - 3450}{25/\sqrt{10}} = 12.65$$

7) Because 12.65 > 1.833, reject H_0.

There is enough evidence to conclude that the mean molecular weight is greater than 3450. However, there is not enough evidence to conclude that it is less than 3550. As a result, we cannot conclude that the new supplier provides a molecular weight that is equivalent to the current one at $\alpha = 0.05$.

9-132

a)

1) The parameter of interest is the mean breaking strength , μ.

2) $H_0 : \mu = 9.5$

3) $H_1 : \mu \geq 9.5$

b)

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,49} = 1.677$ for n = 50 and $\alpha = 0.05$.

6) $\bar{x} = 9.31$, $s = 0.22$, n = 50

$$t_0 = \frac{9.31 - 9.50}{0.22/\sqrt{50}} = -6.107$$

7) Because -6.107 < 1.677, fail to reject H_0.

There is not enough evidence to conclude that the mean breaking strength of the insulators is at least 9.5 psi and that the process by which the insulators are manufactured is equivalent to the standard.

9-133

a)

1) The parameter of interest is the mean bond strength , μ.

2) $H_0 : \mu = 9750$

3) $H_1 : \mu \geq 9750$

b)

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,5} = 2.015$ for n = 6 and $\alpha = 0.05$.

6) $\bar{x} = 9360$, $s = 42.6$, n = 6

$$t_0 = \frac{9360 - 9750}{42.6/\sqrt{6}} = -22.425$$

7) Because -22.425 < 2.015, fail to reject H_0.

There is not enough evidence to conclude that the mean bond strength of cement product is at least 9750 psi and that the process by which the cement product is manufactured is equivalent to the standard.

9-134

$\chi_0^2 = -2[\ln(0.12) + \ln(0.08) + ... + \ln(0.06)] = 46.22$ with $2m = 2(10) = 20$ degrees of freedom.

The *P*-value for this statistic is less than 0.01 (≈ 0.0007). In conclusion, we reject the shared null hypothesis.

9-135

$\chi_0^2 = -2[\ln(0.15) + \ln(0.83) + ... + \ln(0.13)] = 37.40$ with $2m = 2(8) = 16$ degrees of freedom.

The *P*-value for this statistic is less than 0.01 (≈ 0.0018). In conclusion, we reject the shared null hypothesis.

<u>Section 9-11</u>

9-136

$H_0 : \sigma = 0.2$
$H_1 : \sigma < 0.2$

$\chi_0^2 = -2[\ln(0.15) + \ln(0.091) + ... + \ln(0.06)] = 33.66$ with $2m = 2(6) = 12$ degrees of freedom.

The *P*-value for this statistic is less than 0.01 (≈ 0.0007). As a result, we reject the shared null hypothesis. There is sufficient evidence to conclude that the standard deviation of fill volume is less than 0.2 oz.

9-137

$H_0 : \mu = 22$
$H_1 : \mu \neq 22$

$\chi_0^2 = -2[\ln(0.065) + \ln(0.0924) + ... + \ln(0.021)] = 30.57$ with $2m = 2(5) = 10$ degrees of freedom.

The *P*-value for this statistic is less than 0.01 (≈ 0.0006). As a result, we reject the shared null hypothesis. There is sufficient evidence to conclude that the mean package weight is not equal to 22 oz.

Supplemental Exercises

9-138

a) SE Mean $= \dfrac{\sigma}{\sqrt{N}} = \dfrac{1.5}{\sqrt{N}} = 0.401$, so n = 14

$$z_0 = \frac{26.541 - 26}{1.5/\sqrt{14}} = 1.3495$$

P-value = $1 - \Phi(Z_0) = 1 - \Phi(1.3495) = 1 - 0.9114 = 0.0886$

b) A one-sided test because the alternative hypothesis is $\mu > 26$.

c) Because $z_0 < 1.65$ and the P-value = 0.0886 > α = 0.05, we fail to reject the null hypothesis at the 0.05 level of significance.

d) 95% CI of the mean is $\bar{x} - z_{0.025}\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{0.025}\dfrac{\sigma}{\sqrt{n}}$

$$26.541 - (1.96)\frac{1.5}{\sqrt{14}} < \mu < 26.541 + (1.96)\frac{1.5}{\sqrt{14}}$$

$$25.7553 < \mu < 27.3268$$

9-139

a) Degrees of freedom = n – 1 = 16 – 1 = 15.

b) SE Mean $= \dfrac{S}{\sqrt{N}} = \dfrac{4.61}{\sqrt{16}} = 1.1525$

$$t_0 = \frac{98.33 - 100}{4.61/\sqrt{16}} = -1.4490$$

$t_0 = -1.4490$ with df = 15, so 2(0.05) < P-value < 2(0.1). That is, 0.1 < P-value < 0.2.

95% CI of the mean is $\bar{x} - t_{0.025,15}\dfrac{S}{\sqrt{n}} < \mu < \bar{x} + t_{0.025,15}\dfrac{S}{\sqrt{n}}$

$$98.33 - (2.131)\frac{4.61}{\sqrt{16}} < \mu < 98.33 + (2.131)\frac{4.61}{\sqrt{16}}$$

$$95.874 < \mu < 100.786$$

c) Because the P-value > α = 0.05 we fail to reject the null hypothesis at the 0.05 level of significance.

d) $t_{0.05,15} = 1.753$. Because $t_0 = -1.4490 < t_{0.05,15} = 1.753$ we fail to reject the null hypothesis at the 0.05 level of significance.

9-140

a) Degree of freedom = n – 1 = 25 – 1 = 24.

b) SE Mean $= \dfrac{s}{\sqrt{N}} = \dfrac{s}{\sqrt{25}} = 0.631$, so s = 3.155

$$t_0 = \frac{84.331 - 85}{3.155/\sqrt{25}} = -1.06$$

$t_0 = -1.06$ with df = 24, so 0.1 < P-value < 0.25

c) Because the P-value > α = 0.05 we fail to reject the null hypothesis at the 0.05 level of significance.

d) 95% upper CI of the mean is $\mu < \bar{x} + t_{0.05,24}\dfrac{S}{\sqrt{n}}$

$$\mu < 84.331 + (1.711)\frac{3.155}{\sqrt{25}}$$
$$\mu < 85.4106$$

e) If the null hypothesis is changed to mu = 100 versus mu > 100,

$$t_0 = \frac{84.331 - 100}{3.155/\sqrt{25}} = -24.832$$

$t_0 = -24.832$ and $t_{0.05,24} = 1.711$ with df = 24.

Because $t_0 << t_{0.05,24}$ we fail to reject the null hypothesis at the 0.05 level of significance.

9-141

a) The null hypothesis is μ = 12 versus μ > 12

$\bar{x} = 12.4737$, S = 3.6266, and N = 19

$t_0 = \dfrac{12.4737 - 12}{3.6266/\sqrt{19}} = 0.5694$ with df = 19 – 1 = 18.

The P-value falls between two values 0.257 (α = 0.4) and 0.688 (α = 0.25). Thus, 0.25 < P-value < 0.4.
Because the P-value > α = 0.05 we fail to reject the null hypothesis at the 0.05 level of significance.

b) 95% two-sided CI of the mean is $\bar{x} - t_{0.025,18}\dfrac{S}{\sqrt{n}} < \mu < \bar{x} + t_{0.025,18}\dfrac{S}{\sqrt{n}}$

$$12.4737 - (2.101)\frac{3.6266}{\sqrt{19}} < \mu < 12.4737 + (2.101)\frac{3.6266}{\sqrt{19}}$$
$$10.7257 < \mu < 14.2217$$

9-142

a) The null hypothesis is μ = 300 versus μ < 300

$\bar{x} = 275.333$, s = 42.665, and n = 6

$$t_0 = \frac{275.333 - 300}{42.665/\sqrt{6}} = -1.4162$$

and $t_{0.05,5} = 2.015$. Because $t_0 > -t_{0.05,5} = -2.015$ we fail to reject the null hypothesis at the 0.05 level of significance.

b) Yes, because the sample size is very small the central limit theorem's conclusion that the distribution of the sample mean is approximately normally distributed is a concern.

c) 95% two-sided CI of the mean is $\bar{x} - t_{0.025,5}\dfrac{S}{\sqrt{n}} < \mu < \bar{x} + t_{0.025,5}\dfrac{S}{\sqrt{n}}$

$$275.333 - (2.571)\frac{42.665}{\sqrt{6}} < \mu < 275.333 + (2.571)\frac{42.665}{\sqrt{6}}$$
$$230.5515 < \mu < 320.1145$$

9-143 For α = 0.01

a) n = 25 $\beta = \Phi\left(z_{0.01} + \frac{85-86}{16/\sqrt{25}}\right) = \Phi(2.33-0.31) = \Phi(2.02) = 0.9783$

n = 100 $\beta = \Phi\left(z_{0.01} + \frac{85-86}{16/\sqrt{100}}\right) = \Phi(2.33-0.63) = \Phi(1.70) = 0.9554$

n = 400 $\beta = \Phi\left(z_{0.01} + \frac{85-86}{16/\sqrt{400}}\right) = \Phi(2.33-1.25) = \Phi(1.08) = 0.8599$

n = 2500 $\beta = \Phi\left(z_{0.01} + \frac{85-86}{16/\sqrt{2500}}\right) = \Phi(2.33-3.13) = \Phi(-0.80) = 0.2119$

b) n = 25 $z_0 = \frac{86-85}{16/\sqrt{25}} = 0.31$ *P*-value: $1-\Phi(0.31) = 1-0.6217 = 0.3783$

n = 100 $z_0 = \frac{86-85}{16/\sqrt{100}} = 0.63$ *P*-value: $1-\Phi(0.63) = 1-0.7357 = 0.2643$

n = 400 $z_0 = \frac{86-85}{16/\sqrt{400}} = 1.25$ *P*-value: $1-\Phi(1.25) = 1-0.8944 = 0.1056$

n = 2500 $z_0 = \frac{86-85}{16/\sqrt{2500}} = 3.13$ *P*-value: $1-\Phi(3.13) = 1-0.9991 = 0.0009$

The result would be statistically significant when n = 2500 at $\alpha = 0.01$

9-144 Sample Mean = $\hat{p}$ Sample Variance = $\frac{\hat{p}(1-\hat{p})}{n}$

	Sample Size, n	Sampling Distribution	Sample Mean	Sample Variance
a)	50	Normal	p	$\frac{p(1-p)}{50}$
b)	80	Normal	p	$\frac{p(1-p)}{80}$
c)	100	Normal	p	$\frac{p(1-p)}{100}$

d) As the sample size increases, the variance of the sampling distribution decreases.

9-145

	n	Test statistic	P-value	conclusion
a)	50	$z_0 = \frac{0.095-0.10}{\sqrt{0.10(1-0.10)/50}} = -0.12$	0.4522	Fail to reject H_0
b)	100	$z_0 = \frac{0.095-0.10}{\sqrt{0.10(1-0.10)/100}} = -0.15$	0.4404	Fail to reject H_0
c)	500	$z_0 = \frac{0.095-0.10}{\sqrt{0.10(1-0.10)/500}} = -0.37$	0.3557	Fail to reject H_0

d)	1000	$z_0 = \dfrac{0.095 - 0.10}{\sqrt{0.10(1-0.10)/1000}} = -0.53$	0.2981	Fail to reject H_0

e) The P-value decreases as the sample size increases.

9-146 $\sigma = 12$, $\delta = 205 - 200 = 5$, $\frac{\alpha}{2} = 0.025$, $z_{0.025} = 1.96$,

a) n = 20: $\beta = \Phi\left(1.96 - \frac{5\sqrt{20}}{12}\right) = \Phi(0.163) = 0.564$

b) n = 50: $\beta = \Phi\left(1.96 - \frac{5\sqrt{50}}{12}\right) = \Phi(-0.986) = 1 - \Phi(0.986) = 1 - 0.839 = 0.161$

c) n = 100: $\beta = \Phi\left(1.96 - \frac{5\sqrt{100}}{12}\right) = \Phi(-2.207) = 1 - \Phi(2.207) = 1 - 0.9884 = 0.0116$

d) β (probability of a Type II error) decreases as the sample size increases because the variance of the sample mean decreases. Consequently, the probability of observing a sample mean in the acceptance region centered about the incorrect value of 200 ml/h decreases with larger n.

9-147 $\sigma = 14$, $\delta = 205 - 200 = 5$, $\frac{\alpha}{2} = 0.025$, $z_{0.025} = 1.96$,

a) n = 20: $\beta = \Phi\left(1.96 - \frac{5\sqrt{20}}{14}\right) = \Phi(0.362) = 0.6406$

b) n = 50: $\beta = \Phi\left(1.96 - \frac{5\sqrt{50}}{14}\right) = \Phi(-0.565) = 1 - \Phi(0.565) = 1 - 0.7123 = 0.2877$

c) n = 100: $\beta = \Phi\left(1.96 - \frac{5\sqrt{100}}{14}\right) = \Phi(-1.611) = 1 - \Phi(1.611) = 1 - 0.9463 = 0.0537$

d) The probability of a Type II error increases with an increase in the standard deviation.

9-148 $\sigma = 8$, $\delta = 204 - 200 = 4$, $\frac{\alpha}{2} = 0.025$, $z_{0.025} = 1.96$.

a) n = 20: $\beta = \Phi\left(1.96 - \frac{4\sqrt{20}}{8}\right) = \Phi(-0.28) = 1 - \Phi(0.28) = 1 - 0.61026 = 0.38974$

Therefore, power = $1 - \beta = 0.61026$

b) n = 50: $\beta = \Phi\left(1.96 - \frac{4\sqrt{50}}{8}\right) = \Phi(-2.58) = 1 - \Phi(2.58) = 1 - 0.99506 = 0.00494$

Therefore, power = $1 - \beta = 0.995$

c) n = 100: $\beta = \Phi\left(1.96 - \frac{4\sqrt{100}}{8}\right) = \Phi(-3.04) = 1 - \Phi(3.04) = 1 - 0.99882 = 0.00118$

Therefore, power = $1 - \beta = 0.9988$

d) As sample size increases, and all other values are held constant, the power increases because the variance of the sample mean decreases. Consequently, the probability of a Type II error decreases, which implies the power increases.

9-149 a) $\alpha = 0.05$

n = 100 $\beta = \Phi\left(z_{0.05} + \frac{0.5-0.6}{\sqrt{0.5(0.5)/100}}\right) = \Phi(1.65-2.0) = \Phi(-0.35) = 0.3632$

$Power = 1-\beta = 1-0.3632 = 0.6368$

n = 150 $\beta = \Phi\left(z_{0.05} + \frac{0.5-0.6}{\sqrt{0.5(0.5)/100}}\right) = \Phi(1.65-2.45) = \Phi(-0.8) = 0.2119$

$Power = 1-\beta = 1-0.2119 = 0.7881$

n = 300 $\beta = \Phi\left(z_{0.05} + \frac{0.5-0.6}{\sqrt{0.5(0.5)/300}}\right) = \Phi(1.65-3.46) = \Phi(-1.81) = 0.03515$

$Power = 1-\beta = 1-0.03515 = 0.96485$

b) $\alpha = 0.01$

n = 100 $\beta = \Phi\left(z_{0.01} + \frac{0.5-0.6}{\sqrt{0.5(0.5)/100}}\right) = \Phi(2.33-2.0) = \Phi(0.33) = 0.6293$

$Power = 1-\beta = 1-0.6293 = 0.3707$

n = 150 $\beta = \Phi\left(z_{0.01} + \frac{0.5-0.6}{\sqrt{0.5(0.5)/100}}\right) = \Phi(2.33-2.45) = \Phi(-0.12) = 0.4522$

$Power = 1-\beta = 1-0.4522 = 0.5478$

n = 300 $\beta = \Phi\left(z_{0.01} + \frac{0.5-0.6}{\sqrt{0.5(0.5)/300}}\right) = \Phi(2.33-3.46) = \Phi(-1.13) = 0.1292$

$Power = 1-\beta = 1-0.1292 = 0.8702$

Decreasing the value of α decreases the power of the test for the different sample sizes.

c) $\alpha = 0.05$

n = 100 $\beta = \Phi\left(z_{0.05} + \frac{0.5-0.8}{\sqrt{0.5(0.5)/100}}\right) = \Phi(1.65-6.0) = \Phi(-4.35) \cong 0.0$

$Power = 1-\beta = 1-0 \cong 1$

The true value of *p* has a large effect on the power. The greater is the difference of *p* from p_0,the larger is the power of the test.

d)

$$n = \left(\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} - z_\beta \sqrt{p(1-p)}}{p - p_0} \right)^2$$

$$= \left(\frac{2.58\sqrt{0.5(1-0.50)} - 1.65\sqrt{0.6(1-0.6)}}{0.6-0.5} \right)^2 = (4.82)^2 = 23.2 \cong 24$$

$$n = \left(\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} - z_\beta \sqrt{p(1-p)}}{p - p_0} \right)^2$$

$$= \left(\frac{2.58\sqrt{0.5(1-0.50)} - 1.65\sqrt{0.8(1-0.8)}}{0.8-0.5} \right)^2 = (2.1)^2 = 4.41 \cong 5$$

The true value of p has a large effect on the sample size. The greater is the distance of p from p_0,the smaller is the sample size that is required.

9-150 a) Rejecting a null hypothesis provides a *stronger conclusion* than failing to reject a null hypothesis. Therefore, place what we are trying to demonstrate in the alternative hypothesis.

Assume that the data follow a normal distribution.

b)
1) the parameter of interest is the mean weld strength, μ.
2) $H_0 : \mu = 150$
3) $H_1 : \mu > 150$
4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

5) Because no value of α is given, we calculate the P-value
6) $\bar{x} = 153.7$, s= 11.3, n=20

$$t_0 = \frac{153.7 - 150}{11.3/\sqrt{20}} = 1.46$$

P-value = P(t > 1.46) = 0.05 < P-value < 0.10

7) There is some modest evidence to support the claim that the weld strength exceeds 150 psi. If we used α = 0.01 or 0.05, we would fail to reject the null hypothesis, thus the claim would not be supported. If we used α = 0.10, we would reject the null in favor of the alternative and conclude the weld strength exceeds 150 psi.

9-151

a)

$$d = \frac{\delta}{\sigma} = \frac{|\mu - \mu_0|}{\sigma} = \frac{|73 - 75|}{1} = 2$$

Using the OC curve for α = 0.05, d = 2, and n = 10, $\beta \cong 0.0$ and power of $1–0.0 \cong 1$.

$$d = \frac{\delta}{\sigma} = \frac{|\mu - \mu_0|}{\sigma} = \frac{|72 - 75|}{1} = 3$$

Using the OC curve for α = 0.05, d = 3, and n = 10, $\beta \cong 0.0$ and power of $1–0.0 \cong 1$.

b) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|73-75|}{1} = 2$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, $d = 2$, and $\beta \cong 0.1$ (Power=0.9), $n^* = 5$.

Therefore, $n = \dfrac{n^*+1}{2} = \dfrac{5+1}{2} = 3$

$d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|72-75|}{1} = 3$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, $d = 3$, and $\beta \cong 0.1$ (Power=0.9), $n^* = 3$.

Therefore, $n = \dfrac{n^*+1}{2} = \dfrac{3+1}{2} = 2$

c) $\sigma = 2$

$$d = \frac{\delta}{\sigma} = \frac{|\mu - \mu_0|}{\sigma} = \frac{|73-75|}{2} = 1$$

Using the OC curve for $\alpha = 0.05$, $d = 1$, and n = 10, $\beta \cong 0.10$ and power of $1- 0.10 \cong 0.90$.

$$d = \frac{\delta}{\sigma} = \frac{|\mu - \mu_0|}{\sigma} = \frac{|72-75|}{2} = 1.5$$

Using the OC curve for $\alpha = 0.05$, $d = 1.5$, and n = 10, $\beta \cong 0.04$ and power of $1- 0.04 \cong 0.96$.

$$d = \frac{\delta}{\sigma} = \frac{|\mu - \mu_0|}{\sigma} = \frac{|73-75|}{2} = 1$$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, $d = 1$, and $\beta \cong 0.1$ (Power=0.9), $n^* = 10$.

Therefore, $n = \dfrac{n^*+1}{2} = \dfrac{10+1}{2} = 5.5$ $n \cong 6$

$d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|72-75|}{2} = 1.5$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, $d = 3$, and $\beta \cong 0.1$ (Power=0.9), $n^* = 7$.

Therefore, $n = \dfrac{n^*+1}{2} = \dfrac{7+1}{2} = 4$

Increasing the standard deviation decreases the power of the test and increases the sample size required to obtain a certain power.

9-152 Assume the data follow a normal distribution.

a)

1) The parameter of interest is the standard deviation, σ.

2) $H_0 : \sigma^2 = (0.00002)^2$

3) $H_1 : \sigma^2 < (0.00002)^2$

4) The test statistic is: $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma^2}$

5) $\chi^2_{0.99,7} = 1.24$ reject H_0 if $\chi_0^2 < 1.24$ for $\alpha = 0.01$

6) s = 0.00001 and $\alpha = 0.01$

$$\chi_0^2 = \frac{7(0.00001)^2}{(0.00002)^2} = 1.75$$

7) Because 1.75 > 1.24 we fail to reject the null hypothesis. There is insufficient evidence to conclude the standard deviation is at most 0.00002 mm.

b) Although the sample standard deviation is less than the hypothesized value of 0.00002, it is not significantly less (when $\alpha = 0.01$) than 0.00002. The value of 0.00001 could have occurred as a result of sampling variation.

9-153 Assume the data follow a normal distribution.
1) The parameter of interest is the standard deviation of the concentration, σ.
2) $H_0 : \sigma^2 = 4^2$
3) $H_1 : \sigma^2 < 4^2$
4) The test statistic is: $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$
5) Because no value of alpha is specified we calculate the P-value
6) s = 0.004 and n = 10

$$\chi_0^2 = \frac{9(0.004)^2}{(4)^2} = 0.000009$$

P-value = $P\left(\chi^2 < 0.00009\right)$ $\quad P-value \cong 0$

7) Conclusion: The P-value is approximately 0. Therefore we reject the null hypothesis and conclude that the standard deviation of the concentration is less than 4 grams per liter.

9-154

The null hypothesis is that these are 40 observations from a binomial distribution with n = 50 and p must be estimated. Create a table for the number of nonconforming coil springs (value) and the observed frequency. A possible table follows.

Value	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Frequency	0	0	0	1	4	3	4	6	4	3	0	3	3	2	1	1	0	2	1	2

The value of *p* must be estimated. Let the estimate be denoted by $\hat{p}_{\text{sample}}$. The

sample mean $= \dfrac{0(0)+1(0)+2(0)+\cdots+19(2)}{40} = 9.325$ and the mean of a binomial distribution is np.

Therefore,

$$\hat{p}_{sample} = \frac{sample \quad mean}{n} = \frac{9.325}{50} = 0.1865$$

The expected frequencies are obtained from the binomial distribution with n = 50 and p = 0.1865 for 40 observations.

Value	Observed	Expected
0	0	0.001318
1	0	0.015109
2	0	0.084863
3	1	0.311285
4	4	0.838527
5	3	1.768585
6	4	3.040945
7	6	4.382122
8	4	5.399881
9	3	5.777132
10	0	5.43022
11	3	4.526953
12	3	3.372956

13	2	2.260332
14	1	1.369516
15	1	0.753528
16	0	0.377893
17	2	0.173269
18	1	0.072825
19 or more	2	0.042741

Because several of the expected values are less than 3, some cells must be combined resulting in the following table:

Value	Observed	Expected
5 or less	8	3.019686
6	4	3.040945
7	6	4.382122
8	4	5.399881
9	3	5.777132
10	0	5.43022
11	3	4.526953
12	3	3.372956
13 or more	9	5.050105

The degrees of freedom are k – p – 1 = 9 – 1 – 1 = 7

a)
1) Interest is on the form of the distribution for the number of nonconforming coil springs.
2) H_0: The form of the distribution is binomial
3) H_1: The form of the distribution is not binomial
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_0^2 > \chi^2_{0.05,7} = 14.07$ for $\alpha = 0.05$
6)

$$\chi_0^2 = \frac{(8 - 3.019686)^2}{3.019686} + \frac{(4 - 3.040945)^2}{3.040945} + \cdots + \frac{(9 - 5.050105)^2}{5.050105} = 19.888$$

7) Because 19.888 > 14.07 reject H_0. We conclude that the distribution of nonconforming springs is not binomial at α = 0.05.

b) P-value = 0.0058 (from computer software)

9-155

The null hypothesis is that these are 20 observations from a binomial distribution with n = 1000 and p must be estimated. Create a table for the number of errors in a string of 1000 bits (value) and the observed frequency. A possible table follows.

Value	0	1	2	3	4	5
Frequency	3	7	4	5	1	0

The value of *p* must be estimated. Let the estimate be denoted by $\hat{p}_{sample}$. The sample mean $= \dfrac{0(3) + 1(7) + 2(4) + 3(5) + 4(1) + 5(0)}{20} = 1.7$ and the mean of a binomial distribution is np. Therefore,

$$\hat{p}_{sample} = \frac{sample \quad mean}{n} = \frac{1.7}{1000} = 0.0017$$

Value	0	1	2	3	4	5 or more
Observed	3	7	4	5	1	0

Expected	3.64839	6.21282	5.28460	2.99371	1.27067	0.58981

Because several of the expected values are less than 3, some cells are combined resulting in the following table:

Value	0	1	2	3 or more
Observed	3	7	4	6
Expected	3.64839	6.21282	5.28460	4.85419

The degrees of freedom are k – p – 1 = 4 – 1 – 1 = 2

a)
1) Interest is on the form of the distribution for the number of errors in a string of 1000 bits.
2) H_0: The form of the distribution is binomial
3) H_1: The form of the distribution is not binomial
4) The test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

5) Reject H_0 if $\chi_0^2 > \chi_{0.05,2}^2 = 5.99$ for $\alpha = 0.05$
6)

$$\chi_0^2 = \frac{(3-3.64839)^2}{3.64839} + \cdots + \frac{(6-4.85419)^2}{4.85419} = 0.7977$$

7) Because 0.7977 < 9.49 fail to reject H_0. We are unable to reject the null hypothesis that the distribution of the number of errors is binomial at $\alpha = 0.05$.

b) *P*-value = 0.671 (from computer software)

9-156 Divide the real line under a standard normal distribution into eight intervals with equal probability. These intervals are [0, 0.318), [0.318, 0.675), [0.675, 1.150), [1.15, ∞) and their negative counterparts. The probability for each interval is p = 1/8 = 0.125 so the expected cell frequencies are E = *np* = (100) (0.125) = 12.5.

The sample mean and standard deviation are 5421.77 and 33.43616, respectively, and these are used to determine the boundaries of the intervals. That is, the first interval is from negative infinity to 5421.77 + 33.43616(-1.15) = 5383.30. The second interval is from this value to 5421.77 + 33.43616(-0.67449) = 5399.218. The table of ranges and their corresponding frequencies is completed as follows.

Interval	Obs. Frequency.	Exp. Frequency.
$x \le 5383.307$	12	12.5
$5383.307 < x \le 5399.218$	14	12.5
$5399.218 < x \le 5411.116$	12	12.5
$5411.116 < x \le 5421.770$	12	12.5
$5421.770 < x \le 5432.424$	14	12.5
$5432.424 < x \le 5444.322$	9	12.5
$5444.322 < x \le 5460.233$	14	12.5
$x \ge 5460.233$	13	12.5

The test statistic is:

$$\chi_0^2 = \frac{(12-12.5)^2}{12.5} + \frac{(14-12.5)^2}{12.5} + \cdots + \frac{(14-12.5)^2}{12.5} + \frac{(13-12.5)^2}{12.5} = 1.6$$

We reject the null hypothesis if this value exceeds $\chi^2_{0.05,5} = 11.07$. Because $\chi_o^2 < \chi_{0.05,5}^2$, fail to reject the null hypothesis that the data are normally distributed. The P-value = $P(\chi^2 > 1.6) = 0.90$.

9-157 a) In order to use t statistics in hypothesis testing, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true mean concentration of suspended solids, μ.
2) $H_0 : \mu = 50$
3) $H_1 : \mu < 50$

4) Because n >> 30 we can use the normal distribution

$$z_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

5) Reject H_0 if $z_0 < -1.65$ for $\alpha = 0.05$
6) $\bar{x} = 59.87$ s = 12.50 n = 60

$$z_0 = \frac{59.87 - 50}{12.50/\sqrt{60}} = 6.12$$

7) Because 6.12 > –1.65 fail to reject the null hypothesis. There is insufficient evidence to conclude that the true mean concentration of suspended solids is less than 50 ppm at $\alpha = 0.05$.

b) P-value = $\Phi(6.12) \cong 1$

c) We can divide the real line under a standard normal distribution into eight intervals with equal probability. These intervals are [0, 0.32), [0.32, 0.675), [0.675, 1.15), [1.15, ∞) and their negative counterparts. The probability for each interval is p = 1/8 = 0.125 so that the expected cell frequencies are E = np = (60) (0.125) = 7.5.

The sample mean and standard deviation are 59.86667 and 12.49778, respectively, and these are used to determine the boundaries of the intervals. That is, the first interval is from negative infinity to 59.86667 + 12.49778(-1.15) = 45.50. The second interval is from this value to 59.86667 + 12.49778(-0.675) = 51.43. The table of ranges and their corresponding frequencies is completed as follows.

Interval	Obs. Frequency.	Exp. Frequency.
$x \le 45.50$	9	7.5
$45.50 < x \le 51.43$	5	7.5
$51.43 < x \le 55.87$	7	7.5
$55.87 < x \le 59.87$	11	7.5
$59.87 < x \le 63.87$	4	7.5
$63.87 < x \le 68.31$	9	7.5
$68.31 < x \le 74.24$	8	7.5
$x \ge 74.24$	6	7.5

The test statistic is:

$$\chi^2_o = \frac{(9-7.5)^2}{7.5} + \frac{(5-7.5)^2}{7.5} + \cdots + \frac{(8-7.5)^2}{7.5} + \frac{(6-7.5)^2}{7.5} = 5.06$$

and we reject if this value exceeds $\chi^2_{0.05,5} = 11.07$. Because it does not, we fail to reject the hypothesis that the data are normally distributed.

9-158 a) In order to use t statistic in hypothesis testing, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true mean overall distance for this brand of golf ball, μ.
2) $H_0 : \mu = 270$
3) $H_1 : \mu < 270$
4) Since n >> 30 we can use the normal distribution

$$z_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

5) Reject H_0 if $z_0 < -z_\alpha$ where $z_{0.05} = 1.65$ for $\alpha = 0.05$
6) $\bar{x} = 1.25$ s = 0.25 n = 100

$$z_0 = \frac{260.30 - 270.0}{13.41/\sqrt{100}} = -7.23$$

7) Because –7.23 < –1.65 reject the null hypothesis. There is sufficient evidence to indicate that the true mean distance is less than 270 yard at $\alpha = 0.05$.

b) P-value $\cong 0$

c) We can divide the real line under a standard normal distribution into eight intervals with equal probability. These intervals are [0,.32), [0.32, 0.675), [0.675, 1.15), [1.15, ∞) and their negative counterparts. The probability for each interval is p = 1/8 = .125 so the expected cell frequencies are E = np = (100) (0.125) = 12.5.

The sample mean and standard deviation are 260.302 and 13.40828, respectively, and these are used to determine the boundaries of the intervals. That is, the first interval is from negative infinity to 260.302 + 13.40828(-1.15) = 244.88. The second interval is from this value to 260.302 + 13.40828(-0.675) = 251.25. The table of ranges and their corresponding frequencies is completed as follows.

The table of ranges and their corresponding frequencies is completed as follows.

Interval	Obs. Frequency.	Exp. Frequency.
$x \le 244.88$	16	12.5
$244.88 < x \le 251.25$	6	12.5
$251.25 < x \le 256.01$	17	12.5
$256.01 < x \le 260.30$	9	12.5
$260.30 < x \le 264.59$	13	12.5
$264.59 < x \le 269.35$	8	12.5
$269.35 < x \le 275.72$	19	12.5
$x \ge 275.72$	12	12.5

The test statistic is:

$$\chi^2{}_o = \frac{(16-12.5)^2}{12.5} + \frac{(6-12.5)^2}{12.5} + \cdots + \frac{(19-12.5)^2}{12.5} + \frac{(12-12.5)^2}{12.5} = 12$$

and we reject if this value exceeds $\chi^2{}_{0.05,5} = 11.07$. Because it does, we reject the hypothesis that the data are normally distributed.

9-159 a) Assume the data are normally distributed.
1) The parameter of interest is the true mean coefficient of restitution, μ.
2) H_0: $\mu = 0.635$
3) H_1: $\mu > 0.635$
4) Becasue n > 30 we can use the normal distribution

$$z_0 = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

5) Reject H_0 if $z_0 > z_\alpha$ where $z_{0.05}$ =2.33 for $\alpha = 0.01$
6) $\overline{x} = 0.624$ s = 0.0131   n = 40

$$z_0 = \frac{0.624 - 0.635}{0.0131/\sqrt{40}} = -5.31$$

7) Because –5.31< 2.33 fail to reject the null hypothesis. There is insufficient evidence to conclude that the true mean coefficient of restitution is greater than 0.635 at $\alpha = 0.01$.

c) P-value $\Phi(5.31) \cong 1$

d) If the lower bound of the one-sided CI is greater than the value 0.635 then we can conclude that the mean coefficient of restitution is greater than 0.635. Furthermore, a confidence interval provides a range of values for the true mean coefficient that generates information on how much the true mean coefficient differs from 0.635.

9-160 a)
In order to use t statistics in hypothesis testing, we need to assume that the underlying distribution is normal. Use the t-test to test the hypothesis that the true mean is 2.5 mg/L.
1) State the parameter of interest: The parameter of interest is the true mean dissolved oxygen level, μ.
2) State the null hypothesis H_0: $\mu = 2.5$
3) State the alternative hypothesis H_1: $\mu \ne 2.5$
4) Give the statistic

$t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

5) Reject H_0 if $|t_0| < t_{\alpha/2,n-1}$ for $\alpha = 0.05$
6) Sample statistic $\bar{x} = 3.265$ s =2.127 n = 20

$$\text{t-statistic } t_0 = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

7) Draw your conclusion and find the P-value.

b) Assume the data are normally distributed.
1) The parameter of interest is the true mean dissolved oxygen level, μ.
2) H_0: $\mu = 2.5$
3) H_1: $\mu \neq 2.5$
4) Test statistic

$t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$

5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$ where $t_{\alpha/2,n-1} = t_{0.025,19b} = 2.093$ for $\alpha = 0.05$
6) $\bar{x} = 3.265$ s =2.127 n = 20

$$t_0 = \frac{3.265-2.5}{2.127/\sqrt{20}} = 1.608$$

7) Because 1.608 < 2.093 fail to reject the null hypotheses. The sample mean is not significantly different from 2.5 mg/L.

c) The value of 1.608 is found between the columns of 0.05 and 0.1 of Table V. Therefore, 0.1 < P-value < 0.2. Minitab provides a value of 0.124.

d) The confidence interval found in the previous exercise agrees with the hypothesis test above. The value of 2.5 is within the 95% confidence limits. The confidence interval shows that the interval is quite wide due to the large sample standard deviation.

$$\bar{x} - t_{0.025,19}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{0.025,19}\frac{s}{\sqrt{n}}$$

$$3.265 - 2.093\frac{2.127}{\sqrt{20}} \le \mu \le 3.265 + 2.093\frac{2.127}{\sqrt{20}}$$

$$2.270 \le \mu \le 4.260$$

9-161 a)

1) The parameter of interest is the true mean sugar concentration, μ.
2) H_0 : $\mu = 11.5$
3) H_1 : $\mu \neq 11.5$
4) $t_0 = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$
5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$ where $t_{\alpha/2,n-1} = 2.093$ for $\alpha = 0.05$
6) $\bar{x} = 11.47$, s = 0.022 n=20

$$t_0 = \frac{11.47-11.5}{0.022/\sqrt{20}} = -6.10$$

7) Because 6.10 > 2.093 reject the null hypothesis. There is sufficient evidence that the true mean sugar concentration is different from 11.5 at $\alpha = 0.05$.

From Table V the t_0 value in absolute value is greater than the value corresponding to 0.0005 with 19 degrees of freedom. Therefore 2(0.0005) = 0.001 > P-value

b) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|11.4 - 11.5|}{0.022} = 4.54$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, d = 4.54, and n =20 we find $\beta \cong 0$ and Power $\cong 1$.

c) $d = \dfrac{\delta}{\sigma} = \dfrac{|\mu - \mu_0|}{\sigma} = \dfrac{|11.45 - 11.5|}{0.022} = 2.27$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, d = 2.27, and $1 - \beta > 0.9$ ($\beta < 0.1$), we find that n should be at least 5.

d) 95% two sided confidence interval

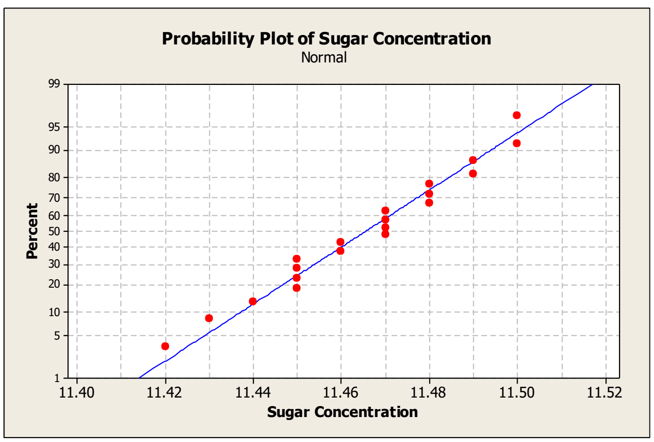
$$\bar{x} - t_{0.025,19}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_{0.025,19}\left(\frac{s}{\sqrt{n}}\right)$$

$$11.47 - 2.093\left(\frac{0.022}{\sqrt{20}}\right) \le \mu \le 11.47 + 2.093\left(\frac{0.022}{\sqrt{20}}\right)$$

$$11.46 \le \mu \le 11.48$$

We conclude that the mean sugar concentration content is not equal to 11.5 because that value is not inside the confidence interval.

e) The normality plot below indicates that the normality assumption is reasonable.



9-162

a) $z_0 = \dfrac{x - np_0}{\sqrt{np_0(1 - p_0)}} = \dfrac{53 - 225(0.25)}{\sqrt{225(0.25)(0.75)}} = -0.5004$

The P-value = $\Phi(-0.5004) = 0.3084$

b) Because the P-value = 0.3084 > α = 0.05 we fail to reject the null hypothesis at the 0.05 level of significance.

c) The normal approximation is appropriate because np > 5 and n(p-1) > 5.

d) $\hat{p} = \dfrac{53}{225} = 0.2356$

The 95% upper confidence interval is:

$$p \le \hat{p} + z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$p \le 0.2356 + 1.65\sqrt{\frac{0.2356(0.7644)}{225}}$$

$$p \le 0.2823$$

e) P-value = $2(1 - \Phi(0.5004)) = 2(1 - 0.6916) = 0.6168$.

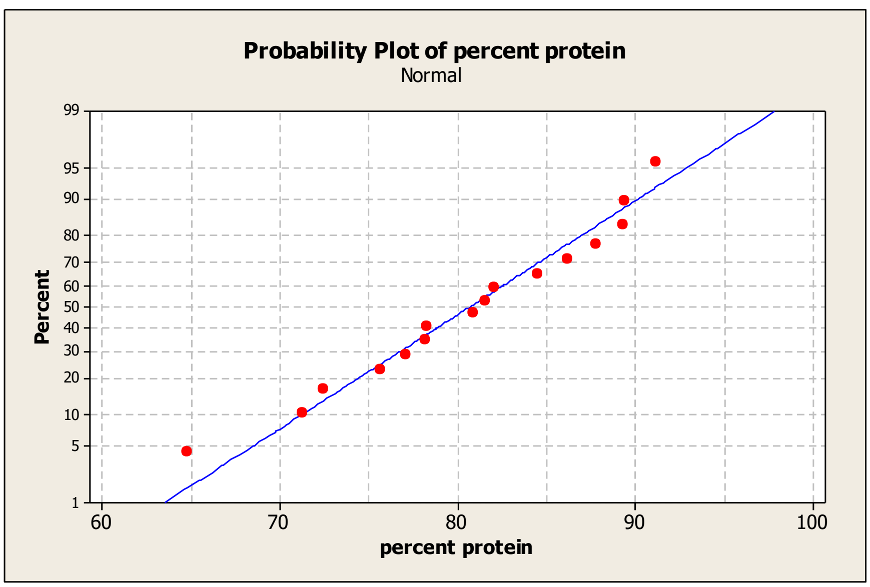
9-163 a)

1) The parameter of interest is the true mean percent protein, μ.
2) $H_0 : \mu = 80$
3) $H_1 : \mu > 80$
4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$
5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,15} = 1.753$ for $\alpha = 0.05$
6) $\bar{x} = 80.68$ $s = 7.38$ $n = 16$

$$t_0 = \frac{80.68 - 80}{7.38/\sqrt{16}} = 0.37$$

7) Because $0.37 < 1.753$ fail to reject the null hypothesis. There is not sufficient evidence to indicate that the true mean percent protein is greater than 80 at $\alpha = 0.05$.

b) From the normal probability plot, the normality assumption seems reasonable:



c) From Table V, $0.25 <$ P-value < 0.4

9-164

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.
1) The parameter of interest is the true variance of tissue assay, σ^2.
2) $H_0 : \sigma^2 = 0.6$
3) $H_1 : \sigma^2 \ne 0.6$

4) $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma^2}$

5) Reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$ where $\alpha = 0.01$ and $\chi^2_{0.995,11} = 2.60$ or $\chi_0^2 > \chi^2_{\alpha/2,n-1}$ where $\alpha = 0.01$ and $\chi^2_{0.005,11} = 26.76$ for n = 12

6) n = 12, s = 0.758

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{11(0.758)^2}{0.6} = 10.53$$

7) Because 2.6 <10.53 < 26.76 we fail to reject H_0. There is not sufficient evidence to conclude the true variance of tissue assay differs from 0.6 at $\alpha = 0.01$.

b) 0.1 < P-value/2 < 0.5, so that 0.2 < P-value < 1

c) 99% confidence interval for σ, first find the confidence interval for σ^2

For $\alpha = 0.05$ and n = 12, $\chi^2_{0.995,11} = 2.60$ and $\chi^2_{0.005,11} = 26.76$

$$\frac{11(0.758)^2}{26.76} \le \sigma^2 \le \frac{11(0.758)^2}{2.60}$$

$$0.236 \le \sigma^2 \le 2.43$$

$$0.486 \le \sigma \le 1.559$$

Because 0.6 falls within the 99% confidence bound there is not sufficient evidence to conclude that the population variance differs from 0.6

9-165

a) In order to use the χ^2 statistic in hypothesis testing and confidence interval construction, we need to assume that the underlying distribution is normal.

1) The parameter of interest is the true variance of the ratio between the numbers of symmetrical and total synapses, σ^2.

2) $H_0: \sigma^2 = 0.02$

3) $H_1: \sigma^2 \ne 0.02$

4) $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma^2}$

5) Reject H_0 if $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.975,30} = 16.79$ or $\chi_0^2 > \chi^2_{\alpha/2,n-1}$ where $\alpha = 0.05$ and $\chi^2_{0.025,30} = 46.98$ for n = 31

6) n = 31, s = 0.198

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{30(0.198)^2}{0.02} = 58.81$$

7) Because 58.81 > 46.98 reject H_0. The true variance of the ratio between the numbers of symmetrical and total synapses is different from 0.02 at $\alpha = 0.05$.

b) P-value/2 < 0.005 so that P-value < 0.01

9-166 a)

1) The parameter of interest is the true mean of cut-on wave length, μ.

2) $H_0: \mu = 6.5$

3) $H_1: \mu \ne 6.5$

4) $t_0 = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

5) Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$. Since no value of α is given, we will assume that $\alpha = 0.05$. So $t_{\alpha/2,n-1} = 2.228$

6) $\bar{x} = 6.55$, s = 0.35 n=11

$$t_0 = \frac{6.55 - 6.5}{0.35/\sqrt{11}} = 0.47$$

7) Because 0.47< 2.228, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that the true mean of cut-on wave length differs from 6.5 at $\alpha = 0.05$.

b) From Table V the t_0 value is found between the values of 0.25 and 0.4 with 10 degrees of freedom, so 0.5 < *P*-value < 0.8

c) $d = \frac{\delta}{\sigma} = \frac{|\mu - \mu_0|}{\sigma} = \frac{|6.25 - 6.5|}{0.35} = 0.71$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, d = 0.71, and $1 - \beta > 0.95$ ($\beta < 0.05$). We find that n should be at least 30.

d) $d = \frac{\delta}{\sigma} = \frac{|\mu - \mu_0|}{\sigma} = \frac{|6.95 - 6.5|}{0.35} = 1.28$

Using the OC curve, Chart VII e) for $\alpha = 0.05$, n =11, d = 1.28, we find $\beta \approx 0.1$

9-167 a)

1) the parameter of interest is the variance of fatty acid measurements, σ^2
2) $H_0 : \sigma^2 = 1.0$
3) $H_1 : \sigma^2 \neq 1.0$
4) The test statistic is: $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$
5) Reject H_0 if $\chi_0^2 < \chi_{0.995,5}^2 = 0.41$ or reject H_0 if $\chi_0^2 > \chi_{0.005,5}^2 = 16.75$ for α=0.01 and n = 6
6) n = 6, s = 0.319

$$\chi_0^2 = \frac{5(0.319)^2}{1^2} = 0.509$$

P-value: 0.005 < P-value/2 < 0.01 so that 0.01 < P-value < 0.02

7) Because the statistic 0.509 > 0.41from the table, fail to reject the null hypothesis at $\alpha = 0.01$. There is insufficient evidence to conclude that the variance differs from 1.0.

b)
1) the parameter of interest is the variance of fatty acid measurements, σ^2 (now n=51)
2) $H_0 : \sigma^2 = 1.0$
3) $H_1 : \sigma^2 \neq 1.0$
4) The test statistic is: $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$
5) Reject H_0 if $\chi_0^2 < \chi_{0.995,50}^2 = 27.99$ or reject H_0 if $\chi_0^2 > \chi_{0.005,50}^2 = 79.49$ for α=0.01 and n = 51

6) n = 51, s = 0.319

$$\chi_0^2 = \frac{50(0.319)^2}{1^2} = 5.09$$

P-value/2 < 0.005 so that P-value < 0.01

7) Because 5.09 < 27.99 reject the null hypothesis. There is sufficient evidence to conclude that the variance differs from1.0 at $\alpha = 0.01$.

c) The sample size changes the conclusion that is drawn. With a small sample size, we fail to reject the null hypothesis. However, a larger sample size allows us to conclude the null hypothesis is false.

9-168 a)

1) the parameter of interest is the standard deviation, σ

2) $H_0 : \sigma^2 = 400$

3) $H_1 : \sigma^2 < 400$

4) The test statistic is: $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2}$

5) No value of α is given, so that no critical value is given. We will calculate the P-value.

6) n = 10, s = 15.7

$$\chi_0^2 = \frac{9(15.7)^2}{400} = 5.546$$

P-value = $P\left(\chi^2 < 5.546\right)$ $\qquad 0.1 < \text{P} - \text{value} < 0.5$

7) The P-value is greater than a common significance level α (such as 0.05). Therefore, we fail to reject the null hypothesis. There is insufficient evidence to support the claim that the standard deviation is less than 20 microamps.

b)

7) n = 51, s = 20

$$\chi_0^2 = \frac{50(15.7)^2}{400} = 30.81$$

P-value = $\text{P}\left(\chi^2 < 30.81\right);\ 0.01 < \text{P} - \text{value} < 0.025$

The P-value is less than 0.05. Therefore, we reject the null hypothesis and conclude that the standard deviation is less than 20 microamps.

a) Increasing the sample size increases the test statistic χ_0^2 and therefore decreases the P-value, providing more evidence against the null hypothesis.

9-169

a)

1) The parameter of interest is the activity level of the active ingredient of new generic drug, μ.

2, 3) The null and alternative hypotheses that must be tested follow ($\delta = 2$):

$H_0 : \mu = 102$ and $H_0 : \mu = 98$

$H_1 : \mu \leq 102$ $\qquad H_1 : \mu \geq 98$

b) Test the first hypothesis ($H_0 : \mu = 102$ vs. $H_1 : \mu \leq 102$):

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

5) Reject H_0 if $t_0 < -t_{\alpha,n-1}$ where $t_{0.05,9} = 1.833$ for n = 10 and $\alpha = 0.05$.

6) $\bar{x} = 96$, $s = 1.5$, n = 10

$$t_0 = \frac{96 - 102}{1.5/\sqrt{10}} = -12.65$$

7) Because -12.65<-1.833, reject H_0.

Test the second hypothesis ($H_0 : \mu = 98$ vs. $H_1 : \mu \geq 98$):

4) The test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,9} = 1.833$ for n = 10 and $\alpha = 0.05$.

6) $\bar{x} = 96$, $s = 1.5$, n = 10

$$t_0 = \frac{96 - 98}{1.5/\sqrt{10}} = -4.22$$

7) Because -4.22 < 1.833, fail to reject H_0.

There is enough evidence to conclude that the mean activity level is less than 102 units. However, there is not enough evidence to conclude that it is greater than 98 units. As a result, we cannot conclude that the new drug is equivalent to the current one according to the activity level of the active ingredient at $\alpha = 0.05$.

9-170

$$\chi_0^2 = -2[\ln(0.15) + \ln(0.06) + ... + \ln(0.13)] = 35.499 \text{ with } 2m = 2(8) = 16 \text{ degrees of freedom.}$$

The *P*-value for this statistic is less than 0.01 (≈ 0.0033). In conclusion, we reject the shared null hypothesis.

Mind Expanding Exercises

9-171 a)

H_0: $\mu = \mu_0$

H_1 $\mu \neq \mu_0$

$P(Z > z_\varepsilon) = \varepsilon$ and $P(Z < -z_{\alpha-\varepsilon}) = (\alpha - \varepsilon)$. Therefore $P(Z > z_\varepsilon$ or $Z < -z_{\alpha-\varepsilon}) = (\alpha - \varepsilon) + \varepsilon = \alpha$

b) $\beta = P(-z_{\alpha-\varepsilon} < Z < z_\varepsilon \mid \mu_0 + \delta)$

9-172 a) Reject H_0 if $z_0 < -z_{\alpha-\varepsilon}$ or $z_0 > z_\varepsilon$

$$P(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \mid \mu = \mu_0) + P(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \mid \mu = \mu_0) \text{P}$$

$$P(z_0 < -z_{\alpha-\varepsilon}) + P(z_0 > z_\varepsilon) = \Phi(-z_{\alpha-\varepsilon}) + 1 - \Phi(z_\varepsilon)$$

$$= ((\alpha - \varepsilon)) + (1 - (1 - \varepsilon)) = \alpha$$

b) $\beta = P(z_\varepsilon \le \bar{X} \le z_\varepsilon$ when $\mu_1 = \mu_0 + d$)

$$\beta = P(-z_{\alpha-\varepsilon} < Z_0 < z_\varepsilon \mid \mu_1 = \mu_0 + \delta)$$

$$\beta = P(-z_{\alpha-\varepsilon} < \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}} < z_\varepsilon \mid \mu_1 = \mu_0 + \delta)$$

$$= P(-z_{\alpha-\varepsilon} - \frac{\delta}{\sqrt{\sigma^2/n}} < Z < z_\varepsilon - \frac{\delta}{\sqrt{\sigma^2/n}})$$

$$= \Phi(z_\varepsilon - \frac{\delta}{\sqrt{\sigma^2/n}}) - \Phi(-z_{\alpha-\varepsilon} - \frac{\delta}{\sqrt{\sigma^2/n}})$$

9-173 1) The parameter of interest is the true mean number of open circuits, λ.

2) $H_0 : \lambda = 2$

3) $H_1 : \lambda > 2$

4) Because n > 30 we can use the normal distribution

$$z_0 = \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}}$$

5) Reject H_0 if $z_0 > z_\alpha$ where $z_{0.05} = 1.65$ for $\alpha = 0.05$

6) $\bar{x}$ = 1038/500=2.076 n = 500

$$z_0 = \frac{2.076-2}{\sqrt{2/500}} = 1.202$$

7) Because 1.202 < 1.65 fail to reject the null hypothesis. There is insufficient evidence to indicate that the true mean number of open circuits is greater than 2 at $\alpha = 0.01$

9-174 a)

1) The parameter of interest is the true standard deviation of the golf ball distance σ.
2) H_0: $\sigma = 10$
3) H_1: $\sigma < 10$
4) Because n > 30 we can use the normal distribution

$$z_0 = \frac{S-\sigma_0}{\sqrt{\sigma_0^2/(2n)}}$$

5) Reject H_0 if $z_0 < z_\alpha$ where $z_{0.05}$ =-1.65 for α=0.05
6) s = 13.41, n = 100

$$z_0 = \frac{13.41-10}{\sqrt{10^2/(200)}} = 4.82$$

7) Because 4.82 > -1.65 fail to reject the null hypothesis. There is insufficient evidence to conclude that the true standard deviation is less than 10 at $\alpha = 0.05$

b) 95% percentile: $\theta = \mu + 1.645\sigma$

95% percentile estimator: $\hat{\theta} = \overline{X} + 1.645S$

From the independence

$$SE(\hat{\theta}) \cong \sqrt{\sigma^2/n + 1.645^2\sigma^2/(2n)}$$

The statistic S can be used as an estimator for σ in the standard error formula.

c)

1) The parameter of interest is the true 95th percentile of the golf ball distance θ.
2) H_0: $\theta = 285$
3) H_1: $\theta < 285$
4) Because n > 30 we can use the normal distribution

$$z_0 = \frac{\hat{\theta}-\theta_0}{\hat{S}E(\hat{\theta})}$$

5) Reject H_0 if $z_0 < -1.65$ for $\alpha = 0.05$
6) $\hat{\theta}$ = 282.36 , s = 13.41, n = 100

$$z_0 = \frac{282.36-285}{\sqrt{13.41^2/100 + 1.645^2 13.41^2/200}} = -1.283$$

7) Because -1.283 > -1.65 fail to reject the null hypothesis. There is not sufficient evidence to conclude that the true θ is less than 285 at $\alpha = 0.05$

9-175 1) The parameter of interest is the parameter of an exponential distribution, λ.
2) $H_0 : \lambda = \lambda_0$
3) $H_1 : \lambda \neq \lambda_0$
4) test statistic

$$\chi_0^2 = \frac{2\lambda\sum_{i=1}^{n} X_i - \lambda_0}{\sqrt{2\lambda\sum_{i=1}^{n} X_i}}$$

5) Reject H$_0$ if $\chi_0^2 > \chi^2_{a/2,2n}$ or $\chi_0^2 < \chi^2_{1-a/2,2n}$ for α = 0.05

6) Compute $2\lambda\sum_{i=1}^{n} X_i$ and plug into

$$\chi_0^2 = \frac{2\lambda\sum_{i=1}^{n} X_i - \lambda_0}{\sqrt{2\lambda\sum_{i=1}^{n} X_i}}$$

7) Draw Conclusions

The one-sided hypotheses below can also be tested with the derived test statistic as follows:
1) H$_0$: $\lambda = \lambda_0$ H$_1$: $\lambda > \lambda_0$
Reject H$_0$ if $\chi_0^2 > \chi^2_{a,2n}$
2) H$_0$: $\lambda = \lambda_0$ H$_1$: $\lambda < \lambda_0$
Reject H$_0$ if $\chi_0^2 < \chi^2_{a,2n}$

CHAPTER 10

Section 10-2

10-1 a)

1) The parameter of interest is the difference in means $\mu_1 - \mu_2$. Note that $\Delta_0 = 0$.

2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2} = -1.96$ or $z_0 > z_{\alpha/2} = 1.96$ for $\alpha = 0.05$

6) $\bar{x}_1 = 4.7$ $\bar{x}_2 = 7.8$

$\sigma_1 = 10$ $\sigma_2 = 5$

$n_1 = 10$ $n_2 = 15$

$$z_0 = \frac{(4.7 - 7.8)}{\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}} = -0.9$$

7) Conclusion: Because $-1.96 < -0.9 < 1.96$, do not reject the null hypothesis. There is not sufficient evidence to conclude that the two means differ at $\alpha = 0.05$.

P-value = $2(1 - \Phi(0.9)) = 2(1 - 0.815950) = 0.368$

b) $$\left(\bar{x}_1 - \bar{x}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(4.7 - 7.8) - 1.96\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}} \leq \mu_1 - \mu_2 \leq (4.7 - 7.8) + 1.96\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}$$

$$-9.79 \leq \mu_1 - \mu_2 \leq 3.59$$

With 95% confidence, the true difference in the means is between −9.79 and 3.59. Because zero is contained in this interval, we conclude there is no significant difference between the means. We fail to reject the null hypothesis.

c)

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

$$= \Phi\left(1.96 - \frac{3}{\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}} \right) - \Phi\left(-1.96 - \frac{3}{\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}} \right) = \Phi(1.08) - \Phi(-2.83) = 0.8599 - 0.0023 = 0.86$$

Power = 1 − 0.86 = 0.14

d) Assume the sample sizes are to be equal, use $\alpha = 0.05$, $\beta = 0.05$, and $\delta = 3$

$$n \cong \frac{\left(z_{\alpha/2} + z_\beta\right)^2\left(\sigma_1^2 + \sigma_2^2\right)}{\delta^2} = \frac{(1.96 + 1.645)^2\left(10^2 + 5^2\right)}{(3)^2} = 180.5$$

Use $n_1 = n_2 = 181$

10-2 a)

1) The parameter of interest is the difference in means $\mu_1 - \mu_2$. Note that $\Delta_0 = 0$.

2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_\alpha = -1.645$ for $\alpha = 0.05$

6) $\bar{x}_1 = 14.2$ $\bar{x}_2 = 19.7$

$\sigma_1 = 10$ $\sigma_2 = 5$

$n_1 = 10$ $n_2 = 15$

$$z_0 = \frac{(14.2 - 19.7)}{\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}} = -1.61$$

7) Conclusion: Because –1.61 > -1.645, do not reject the null hypothesis. There is not sufficient evidence to conclude that the two means differ at $\alpha = 0.05$.

P-value = $\Phi(-1.61) = 0.0537$

b) $$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\mu_1 - \mu_2 \le (14.2 - 19.7) + 1.645\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}$$

$$\mu_1 - \mu_2 \le 0.12$$

With 95% confidence, the true difference in the means is less than 0.12. Because zero is contained in this interval, we fail to reject the null hypothesis.

c)

$$\beta = 1 - \Phi\left(-z_\alpha - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

$$= 1 - \Phi\left(-1.65 - \frac{-4}{\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}} \right) = 1 - \Phi(-0.4789) = 0.316$$

Power = 1 – 0.316 = 0.684

d) Assume the sample sizes are to be equal, use $\alpha = 0.05$, $\beta = 0.05$, and $\delta = \Delta - \Delta_0 = 4$

$$n \cong \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2} = \frac{(1.645 + 1.645)^2(10^2 + 5^2)}{(4)^2} = 85$$

Use $n_1 = n_2 = 85$

10-3 a)

1) The parameter of interest is the difference in means $\mu_1 - \mu_2$. Note that $\Delta_0 = 0$.

2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 > 0$ or $\mu_1 > \mu_2$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 > z_\alpha = 2.325$ for $\alpha = 0.01$
6) $\bar{x}_1 = 24.5$ $\bar{x}_2 = 21.3$
$\sigma_1 = 10$ $\sigma_2 = 5$
$n_1 = 10$ $n_2 = 15$

$$z_0 = \frac{(24.5 - 21.3)}{\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}} = 0.937$$

7) Conclusion: Because 0.937 < 2.325, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that the two means differ at $\alpha = 0.01$.

P-value = $1 - \Phi(0.94) = 1 - 0.8264 = 0.1736$

b) $\mu_1 - \mu_2 \ge (\bar{x}_1 - \bar{x}_2) - z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$\mu_1 - \mu_2 \ge (24.5 - 21.3) - 2.325\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}$$

$$\mu_1 - \mu_2 \ge -4.74$$

The true difference in the means is greater than -4.74 with 99% confidence. Because zero is contained in this interval, we fail to reject the null hypothesis.

c)

$$\beta = \Phi\left(z_\alpha - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) = \Phi\left(2.325 - \frac{2}{\sqrt{\frac{(10)^2}{10} + \frac{(5)^2}{15}}} \right) = \Phi(1.74) = 0.959$$

Power = 1 – 0.96 = 0.04

d) Assume the sample sizes are to be equal, use $\alpha = 0.05$, $\beta = 0.05$, and $\Delta = 3$

$$n \cong \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2} = \frac{(1.645 + 1.645)^2(10^2 + 5^2)}{(2)^2} = 339$$

Use $n_1 = n_2 = 339$

10-4 a)
1) The parameter of interest is the difference in fill volume $\mu_1 - \mu_2$. Note that $\Delta_0 = 0$.
2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) H_1 : $\mu_1 - \mu_2 \ne 0$ or $\mu_1 \ne \mu_2$
4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2} = -1.96$ or $z_0 > z_{\alpha/2} = 1.96$ for $\alpha = 0.05$
6) $\bar{x}_1 = 16.014$ $\bar{x}_2 = 16.006$
$\sigma_1 = 0.020$ $\sigma_2 = 0.025$

$n_1 = 10$ $n_2 = 10$

$$z_0 = \frac{(16.014 - 16.006)}{\sqrt{\frac{(0.020)^2}{10} + \frac{(0.025)^2}{10}}} = 0.79$$

7) Conclusion: Because $-1.96 < 0.79 < 1.96$, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the two machine fill volumes differ at $\alpha = 0.05$.

P-value $= 2(1 - \Phi(0.79)) = 2(1 - 0.7852) = 0.429$

b) $\left(\bar{x}_1 - \bar{x}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$(16.014 - 16.006) - 1.96\sqrt{\frac{(0.02)^2}{10} + \frac{(0.025)^2}{10}} \le \mu_1 - \mu_2 \le (16.014 - 16.006) + 1.96\sqrt{\frac{(0.02)^2}{10} + \frac{(0.025)^2}{10}}$$

$$-0.0168 \le \mu_1 - \mu_2 \le 0.0328$$

With 95% confidence, the true difference in the mean fill volumes is between -0.0098 and 0.0298. Because zero is contained in this interval, there is no significant difference between the means.

c)

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right)$$

$$= \Phi\left(1.96 - \frac{0.04}{\sqrt{\frac{(0.032)^2}{10} + \frac{(0.024)^2}{10}}}\right) - \Phi\left(-1.96 - \frac{0.04}{\sqrt{\frac{(0.032)^2}{10} + \frac{(0.024)^2}{10}}}\right)$$

$= \Phi(1.96 - 3.16) - \Phi(-1.96 - 3.16) = \Phi(-1.2) - \Phi(-5.12) = 0.1151 - 0 = 0.1151$

Power $= 1 - 0.1151 = 0.8849$

d) Assume the sample sizes are to be equal, use $\alpha = 0.05$, $\beta = 0.05$, and $\Delta = 0.04$

$$n \cong \frac{\left(z_{\alpha/2} + z_\beta\right)^2\left(\sigma_1^2 + \sigma_2^2\right)}{\delta^2} = \frac{(1.96 + 1.645)^2\left(0.032^2 + 0.024^2\right)}{(0.04)^2} = 12.96$$

Use $n_1 = n_2 = 13$

10-5 a)

1) The parameter of interest is the difference in breaking strengths $\mu_1 - \mu_2$ and $\Delta_0 = 10$

2) $H_0: \mu_1 - \mu_2 = 10$

3) $H_1: \mu_1 - \mu_2 > 10$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 > z_\alpha = 1.645$ for $\alpha = 0.05$

6) $\bar{x}_1 = 162.5$ $\bar{x}_2 = 155.0$ $\delta = 10$

$\sigma_1 = 1.0$ $\sigma_2 = 1.0$

$n_1 = 10$ $n_2 = 12$

$$z_0 = \frac{(162.5 - 155.0) - 10}{\sqrt{\frac{(1.0)^2}{10} + \frac{(1.0)^2}{12}}} = -5.84$$

7) Conclusion: Because $-5.84 < 1.645$ fail to reject the null hypothesis. There is insufficient evidence to support the use of plastic 1 at $\alpha = 0.05$.

P-value= $1-\Phi(-5.84)$ = 1-0=1

b) $\mu_1-\mu_2 \ge (\bar{x}_1-\bar{x}_2)-z_\alpha\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$

$\mu_1-\mu_2 \ge (162.5-155)-1.645\sqrt{\dfrac{(1)^2}{10}+\dfrac{(1)^2}{12}}$

$\mu_1-\mu_2 \ge 6.8$

c)

β = $\Phi\left(1.645-\dfrac{(12-10)}{\sqrt{\dfrac{1}{10}+\dfrac{1}{12}}}\right)=\Phi(-3.03)=0.0012$

Power = 1 – 0.0012 = 0.9988

d)

$$n=\frac{(z_{\alpha/2}+z_\beta)^2(\sigma_1^2+\sigma_2^2)}{(\Delta-\Delta_0)^2}=\frac{(1.645+1.645)^2(1+1)}{(12-10)^2}=5.42\cong 6$$

Yes, the sample size is adequate

10-6 a)

1) The parameter of interest is the difference in mean burning rate, $\mu_1-\mu_2$

2) $H_0: \mu_1-\mu_2=0$ or $\mu_1=\mu_2$

3) $H_1: \mu_1-\mu_2\neq 0$ or $\mu_1\neq\mu_2$

4) The test statistic is

$$z_0=\frac{(\bar{x}_1-\bar{x}_2)-\Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2} = -1.96$ or $z_0 > z_{\alpha/2} = 1.96$ for $\alpha = 0.05$

6) $\bar{x}_1=18 \quad \bar{x}_2=24$

$\sigma_1=3 \quad \sigma_2=3$

$n_1 = 20 \quad n_2 = 20$

$$z_0=\frac{(18-24)}{\sqrt{\dfrac{(3)^2}{20}+\dfrac{(3)^2}{20}}}=-6.32$$

7) Conclusion: Because –6.32 < –1.96 reject the null hypothesis and conclude the mean burning rates differ significantly at $\alpha = 0.05$.

P-value = $2(1-\Phi(6.32))=2(1-1)=0$

b) $(\bar{x}_1-\bar{x}_2)-z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}\le\mu_1-\mu_2\le(\bar{x}_1-\bar{x}_2)+z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$

$(18-24)-1.96\sqrt{\dfrac{(3)^2}{20}+\dfrac{(3)^2}{20}}\le\mu_1-\mu_2\le(18-24)+1.96\sqrt{\dfrac{(3)^2}{20}+\dfrac{(3)^2}{20}}$

$-7.86\le\mu_1-\mu_2\le-4.14$

We are 95% confident that the mean burning rate for solid fuel propellant 2 exceeds that of propellant 1 by between 4.14 and 7.86 cm/s.

c) $\beta = \Phi\left(z_{\alpha/2} - \dfrac{\Delta \ - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \right) - \Phi\left(-z_{\alpha/2} - \dfrac{\Delta - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \right)$

$$= \Phi\left(1.96 - \frac{2.5}{\sqrt{\frac{(3)^2}{20} + \frac{(3)^2}{20}}} \right) - \Phi\left(-1.96 - \frac{2.5}{\sqrt{\frac{(3)^2}{20} + \frac{(3)^2}{20}}} \right)$$

$$= \Phi(1.96 - 2.64) - \Phi(-1.96 - 2.64) = \Phi(-0.68) - \Phi(-4.6) = 0.24825 - 0 = 0.24825$$

d) Assume the sample sizes are to be equal, use $\alpha = 0.05$, β = 1-power=0.1, and $\Delta = 4$

$n \cong \dfrac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2} = \dfrac{(1.96 + 1.28)^2(3^2 + 3^2)}{(4)^2} = 12$

Use $n_1 = n_2 = 12$

10-7 $\bar{x}_1 = 89.6 \quad \bar{x}_2 = 92.5$

$\sigma_1^2 = 1.5 \quad \sigma_2^2 = 1.2$

$n_1 = 15 \quad n_2 = 20$

a)

1) The parameter of interest is the difference in mean road octane number $\mu_1 - \mu_2$ and $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_\alpha = -1.645$ for $\alpha = 0.05$

6) $\bar{x}_1 = 89.6 \quad \bar{x}_2 = 92.5$

$\sigma_1^2 = 1.5 \quad \sigma_2^2 = 1.2$

$n_1 = 15 \quad n_2 = 20$

$$z_0 = \frac{(89.6 - 92.5)}{\sqrt{\frac{1.5}{15} + \frac{1.2}{20}}} = -7.25$$

7) Conclusion: Because $-7.25 < -1.645$ reject the null hypothesis and conclude the mean road octane number for formulation 2 exceeds that of formulation 1 using $\alpha = 0.05$.

P-value $\cong P(z \le -7.25) = 1 - P(z \le 7.25) = 1 - 1 \cong 0$

b) 95% confidence interval:

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(89.6 - 92.5) - 1.96\sqrt{\frac{1.5}{15} + \frac{1.2}{20}} \le \mu_1 - \mu_2 \le (89.6 - 92.5) + 1.96\sqrt{\frac{1.5}{15} + \frac{1.2}{20}}$$

$$-3.684 \le \mu_1 - \mu_2 \le -2.116$$

With 95% confidence, the mean road octane number for formulation 2 exceeds that of formulation 1 by between 2.116 and 3.684.

c) 95% level of confidence, E = 1, and $z_{0.025}$ =1.96

$$n \cong \left(\frac{z_{0.025}}{E}\right)^2 \left(\sigma_1^2 + \sigma_2^2\right) = \left(\frac{1.96}{1}\right)^2 (1.5 + 1.2) = 10.37,$$

Use $n_1 = n_2 = 11$

10-8 a)

1) The parameter of interest is the difference in mean batch viscosity before and after the process change, $\mu_1 - \mu_2$

2) $H_0: \mu_1 - \mu_2 = 10$

3) $H_1: \mu_1 - \mu_2 < 10$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_\alpha$ where $z_{0.1} = -1.28$ for $\alpha = 0.10$

6) $\bar{x}_1 = 750.2$ $\bar{x}_2 = 756.88$ $\Delta_0 = 10$

$\sigma_1 = 20$ $\sigma_2 = 20$

$n_1 = 15$ $n_2 = 8$

$$z_0 = \frac{(750.2 - 756.88) - 10}{\sqrt{\frac{(20)^2}{15} + \frac{(20)^2}{8}}} = -1.90$$

7) Conclusion: Because $-1.90 < -1.28$ reject the null hypothesis and conclude the process change has increased the mean by less than 10.

P-value = $P(Z \le -1.90) = 1 - P(Z \le 1.90) = 1 - 0.97128 = 0.02872$

b)

Case 1: Before Process Change	Case 2: After Process Change
μ_1 = mean batch viscosity before change	μ_2 = mean batch viscosity after change
$\bar{x}_1 = 750.2$	$\bar{x}_2 = 756.88$
$\sigma_1 = 20$	$\sigma_2 = 20$
$n_1 = 15$	$n_2 = 8$

90% confidence on $\mu_1 - \mu_2$, the difference in mean batch viscosity before and after process change:

$$\left(\bar{x}_1 - \bar{x}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(750.2 - 756.88) - 1.645\sqrt{\frac{(20)^2}{15} + \frac{(20)^2}{8}} \le \mu_1 - \mu_2 \le \left(750.2 - 756/88\right) + 1.645\sqrt{\frac{(20)^2}{15} + \frac{(20)^2}{8}}$$

$$-21.08 \le \mu_1 - \mu_2 \le 7.72$$

We are 90% confident that the difference in mean batch viscosity before and after the process change lies within −21.08 and 7.72. Because zero is contained in this interval, we fail to detect a difference in the mean batch viscosity from the process change.

c) Parts (a) and (b) conclude that the mean batch viscosity change is less than 10. This conclusion is obtained from the confidence interval because the interval does not contain the value 10. The upper endpoint of the confidence interval is only 7.72.

10-9 <u>Catalyst 1</u> <u>Catalyst 2</u>

$\bar{x}_1 = 65.22$ $\qquad\qquad$ $\bar{x}_2 = 68.42$

$\sigma_1 = 3$ $\qquad\qquad$ $\sigma_2 = 3$

$n_1 = 10$ $\qquad\qquad$ $n_2 = 10$

a) 95% confidence interval on $\mu_1 - \mu_2$, the difference in mean active concentration

$$\left(\bar{x}_1 - \bar{x}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(65.22 - 68.42) - 1.96\sqrt{\frac{(3)^2}{10} + \frac{(3)^2}{10}} \le \mu_1 - \mu_2 \le \left(65.22 - 68.42\right) + 1.96\sqrt{\frac{(3)^2}{10} + \frac{(3)^2}{10}}$$

$$-5.83 \le \mu_1 - \mu_2 \le -0.57$$

We are 95% confident that the mean active concentration of catalyst 2 exceeds that of catalyst 1 by between 0.57 and 5.83 g/l.

P-value:

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(65.22 - 68.42)}{\sqrt{\frac{3^2}{10} + \frac{3^2}{10}}} = -2.38$$

Then P-value = 2(0.008656) = 0.0173

b) Yes, because the 95% confidence interval does not contain the value zero. We conclude that the mean active concentration depends on the choice of catalyst.

c)

$$\beta = \Phi\left(1.96 - \frac{(5)}{\sqrt{\frac{3^2}{10} + \frac{3^2}{10}}}\right) - \Phi\left(-1.96 - \frac{(5)}{\sqrt{\frac{3^2}{10} + \frac{3^2}{10}}}\right)$$

$$= \Phi(-1.77) - \Phi(-5.69) = 0.038364 - 0$$

$$= 0.038364$$

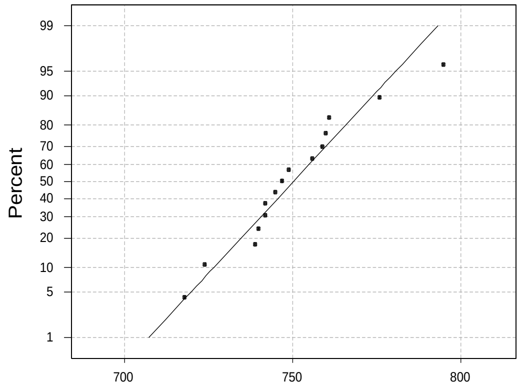
Power = 1 – β = 1– 0.038364 = 0.9616.

d) Calculate the value of *n* using α and β.

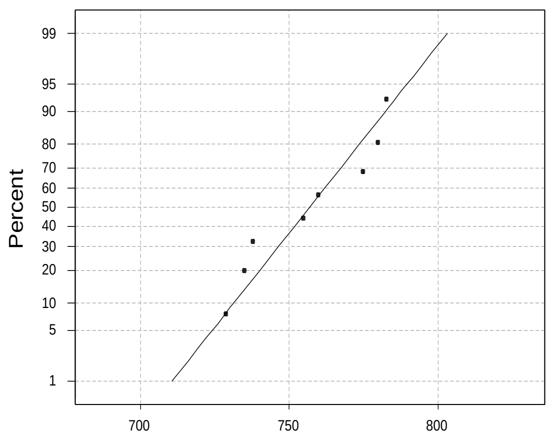
$$n \cong \frac{\left(z_{\alpha/2} + z_\beta\right)^2\left(\sigma_1^2 + \sigma_2^2\right)}{\left(\Delta - \Delta_0\right)^2} = \frac{(1.96 + 1.77)^2(9 + 9)}{(5)^2} = 10.02,$$

Therefore, 10 is only slightly too few samples. The sample sizes are adequate to detect the difference of 5.

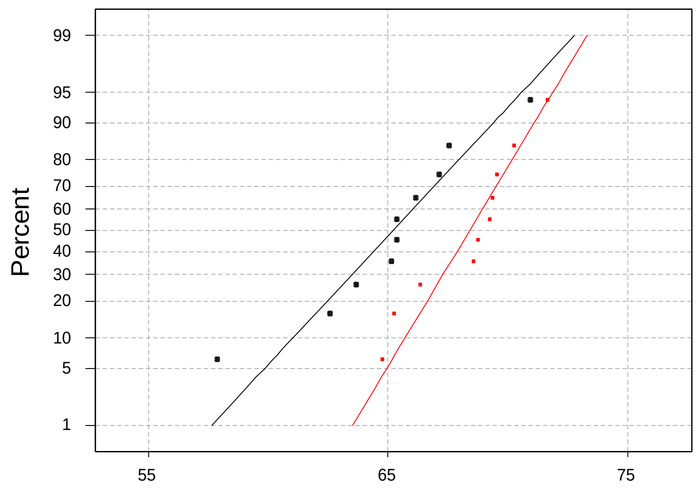
The data from the first sample *n* = 15 appear to be normally distributed.



The data from the second sample n = 8 appear to be normally distributed



Plots for both samples are shown in the following figure.



10-10 1) The parameter of interest is the difference in means $\mu_1 - \mu_2$. Note that $\Delta_0 = 0$.

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2} = -1.96$ or $z_0 > z_{\alpha/2} = 1.96$ for $\alpha = 0.05$

6) $\bar{x}_1 = 6 \quad \bar{x}_2 = 6.2$

$\sigma_1 = 1.5 \qquad \sigma_2 = 0.3$

$n_1 = 6 \qquad n_2 = 6$

$$z_0 = \frac{(6 - 6.2)}{\sqrt{\frac{(1.5)^2}{6} + \frac{(0.3)^2}{6}}} = -0.32036$$

7) Conclusion: Because $-1.96 < -0.32036 < 1.96$, do not reject the null hypothesis. There is not sufficient evidence to conclude that the two means differ at $\alpha = 0.05$.

10-11

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(6 - 6.2) - 1.96\sqrt{\frac{(1.5)^2}{6} + \frac{(0.3)^2}{6}} \le \mu_1 - \mu_2 \le (6 - 6.2) + 1.96\sqrt{\frac{(1.5)^2}{6} + \frac{(0.3)^2}{6}}$$

$$-1.42 \le \mu_1 - \mu_2 \le 1.02$$

With 95% confidence, the true difference in the means is between -1.42 and 1.02. Because zero is contained in this interval, we conclude there is no significant difference between the means. We fail to reject the null hypothesis.

10-12

The expression for probability of type II error β for a two sided alternative hypothesis is

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right)$$

$$\Phi\left(1.96 - \frac{0.25}{\sqrt{\frac{(1.5)^2}{6} + \frac{(0.3)^2}{6}}} \right) - \Phi\left(-1.96 - \frac{0.25}{\sqrt{\frac{(1.5)^2}{6} + \frac{(0.3)^2}{6}}} \right)$$

$= \Phi(1.56) - \Phi(-2.36)$ = 0.9406 – 0.0091 = 0.9315

Power = 1 – 0.9315 = 0.0685

Assume the sample sizes are to be equal, use α = 0.05, β = 0.0685, and

$$n \cong \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2} = \frac{(1.96 + 1.487)^2(1.5^2 + 0.3^2)}{(0.25)^2} = 444.9$$

Use $n_1 = n_2 = 445$

10-13 1) The parameter of interest is the difference in means $\mu_1 - \mu_2$ and $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_\alpha = -1.645$ for $\alpha = 0.05$

6) $\bar{x}_1 = 190$ $\bar{x}_2 = 310$

$\sigma_1 = 15$ $\sigma_2 = 50$

$n_1 = 10$ $n_2 = 4$

$$z_0 = \frac{(190 - 310)}{\sqrt{\frac{15^2}{10} + \frac{50^2}{4}}} = -4.7159$$

7) Conclusion: Because –4.7159< –1.645, reject the null hypothesis and conclude that the mean absorbency of the towels from process 2 exceeds that of process 1 using α = 0.05.

The probability of type I error that is appropriate depends on the relationships between the processes. For example, if an expansive change is planned for process 2, one would not want to incorrectly conclude that is produces a better product.

Section 10-2

10-14 a) $\bar{x}_1 = 10.94$ $\bar{x}_2 = 12.15$ $s_1^2 = 1.26^2$ $s_2^2 = 1.99^2$ $n_1 = 12$ $n_2 = 16$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(12-1)1.26^2 + (16-1)1.99^2}{12 + 16 - 2}} = 1.7194$$

Degree of freedom = $n_1 + n_2 - 2$ = 12 + 16 – 2 = 26.

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(-1.21)}{1.7194\sqrt{\frac{1}{12} + \frac{1}{16}}} = -1.8428$$

P-value = 2[P(t > 1.8428)] and 2(0.025) < P-value < 2(0.05) = 0.05 < P-value < 0.1

This is a two-sided test because the hypotheses are mu1 – mu2 = 0 versus not equal to 0.

b) Because 0.05 < P-value < 0.1 the P-value is greater than α = 0.05. Therefore, we fail to reject the null hypothesis of $\mu_1 - \mu_2 = 0$ at the 0.05 and 0.01 levels of significance.

c) Yes, the sample standard deviations are somewhat different, but not excessively different. Consequently, the assumption that the two population variances are equal is reasonable.

d) P-value = P (t < -1.8428) and 0.025 < P-value < 0.05
Because 0.025 < P-value < 0.05, the P-value is less than $\alpha = 0.05$. Therefore, we reject the null hypothesis of $\mu_1 - \mu_2 = 0$ at the 0.05 level of significance.

10-15 a) $\bar{x}_1 = 54.73$ $\bar{x}_2 = 58.64$ $s_1^2 = 2.13^2$ $s_2^2 = 5.28^2$ $n_1 = 15$ $n_2 = 20$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}} = \frac{\left(\frac{2.13^2}{15} + \frac{5.28^2}{20}\right)^2}{\frac{\left(2.13^2/15\right)^2}{15-1} + \frac{\left(5.28^2/20\right)^2}{20-1}} = 26.45 \approx 26 \text{ (truncated)}$$

The 95% upper one-sided confidence interval: $t_{0.05,26} = 1.706$

$$\mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\mu_1 - \mu_2 \le (54.74 - 58.64) + 1.706\sqrt{\frac{(2.13)^2}{15} + \frac{(5.28)^2}{20}}$$

$$\mu_1 - \mu_2 \le -1.6880$$

P-value = P(t < –3.00): 0.0025 < P-value < 0.005

This is one-sided test because the hypotheses are mu1 – mu2 = 0 versus less than 0.

b) Because 0.0025 < P-value < 0.005 the P-value < $\alpha = 0.05$. Therefore, we reject the null hypothesis of mu1 – mu2 = 0 at the 0.05 or the 0.01 level of significance.

c) Yes, the sample standard deviations are quite different. Consequently, one would not want to assume that the population variances are equal.

d) If the alternative hypothesis were changed to mu1 – mu2 ≠ 0, then the P-value = 2P (t < –3.00) and 0.005 < P-value < 0.01. Because the P-value < $\alpha = 0.05$, we reject the null hypothesis of mu1 – mu2 = 0 at the 0.05 level of significance.

10-16 a) 1) The parameter of interest is the difference in mean, $\mu_1 - \mu_2$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,28} = -2.048$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,28} = 2.048$ for $\alpha = 0.05$

6) $\bar{x}_1 = 4.7$ $\bar{x}_2 = 7.8$ $$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$s_1^2 = 4 \qquad s_2^2 = 6.25$ $\qquad = \sqrt{\dfrac{14(4)+14(6.25)}{28}} = 2.26$

$n_1 = 15 \qquad n_2 = 15$

$$t_0 = \frac{(4.7-7.8)}{2.26\sqrt{\frac{1}{15}+\frac{1}{15}}} = -3.75$$

7) Conclusion: Because –3.75 <-2.048, reject the null hypothesis at $\alpha = 0.05$.

P-value = $2P\left(t > 3.75\right) < 2(0.0005)$, P-value <.001

b) 95% confidence interval: $t_{0.025,28} = 2.048$

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(4.7-7.8) - 2.048(2.26)\sqrt{\frac{1}{15}+\frac{1}{15}} \le \mu_1 - \mu_2 \le (4.7-7.8) + 2.048(2.26)\sqrt{\frac{1}{15}+\frac{1}{15}}$$

$$-4.79 \le \mu_1 - \mu_2 \le -1.41$$

Because zero is not contained in this interval, we reject the null hypothesis that the means are equal.

c) $\Delta = 3$ Use s_p as an estimate of σ.

d = $\dfrac{\mu_2 - \mu_1}{2s_p} = \dfrac{3}{2(2.26)} = 0.66$

Using Chart VII (e) with d = 0.66 and $n = n_1 = n_2$ we obtain $n^* = 2n - 1 = 29$ and $\alpha = 0.05$. Therefore, $\beta = 0.1$ and the power is $1 - \beta = 0.9$

d) $\beta = 0.05$, d= $\dfrac{2}{2(2.26)}$ = 0.44, therefore $n^* \cong 75$ then $n = \dfrac{n^*+1}{2}$ = 38, then $n = n_1 = n_2$ =38

10-17 a)

1) The parameter of interest is the difference in means, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha,n_1+n_2-2}$ where $-t_{0.05,28} = -1.701$ for $\alpha = 0.05$

6)

$\bar{x}_1 = 6.2 \qquad \bar{x}_2 = 7.8$ $\qquad s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

$s_1^2 = 4 \qquad s_2^2 = 6.25$ $\qquad = \sqrt{\dfrac{14(4)+14(6.25)}{28}} = 2.26$

$n_1 = 15 \qquad n_2 = 15$

$$t_0 = \frac{(6.2-7.8)}{2.26\sqrt{\frac{1}{15}+\frac{1}{15}}} = -1.94$$

7) Conclusion: Because -1.94 < –1.701 reject the null hypothesis at the 0.05 level of significance.

P-value = P$(t > 1.94)$ 0.025 < *P*-value < 0.05

b) 95% confidence interval: $t_{0.05,28} = 1.701$

$$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$\mu_1 - \mu_2 \le (6.2-7.8) + 1.701(2.26)\sqrt{\frac{1}{15}+\frac{1}{15}}$$

$$\mu_1 - \mu_2 \le -0.196$$

Because zero is not contained in this interval, we reject the null hypothesis.

c) $\Delta = 3$ Use s_p as an estimate of σ.

d = $\frac{\mu_2-\mu_1}{2s_p} = \frac{3}{2(2.26)} = 0.66$

Using Chart VII (g) with d = 0.66 and $n = n_1 = n_2$ we get $n^* = 2n - 1 = 29$ and α = 0.05 . Therefore, β = 0.05 and the power is 1-β = 0.95

d) β = 0.05, d = $\frac{2.5}{2(2.26)}$ = 0.55. Therefore $n^* \cong 40$ and $n = \frac{n^*+1}{2} \cong 21$. Thus, $n = n_1 = n_2 = 21$

10-18 a)

1) The parameter of interest is the difference in means, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 > 0$ or $\mu_1 > \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 > t_{\alpha,n_1+n_2-2}$ where $t_{0.05,18} = 1.734$ for α = 0.05

6)

$\bar{x}_1 = 7.8$ $\bar{x}_2 = 5.6$

$s_1^2 = 4$ $s_2^2 = 9$

$n_1 = 10$ $n_2 = 10$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{9(4)+9(9)}{18}} = 2.55$$

$$t_0 = \frac{(7.8-5.6)}{2.55\sqrt{\frac{1}{10}+\frac{1}{10}}} = 1.93$$

7) Conclusion: Because 1.93 > 1.714 reject the null hypothesis at the 0.05 level of significance.

P-value = $P(t > 1.93)$ and 0.025 < *P*-value < 0.05

b) 95% confidence interval:

$$\mu_1 - \mu_2 \geq (\bar{x}_1 - \bar{x}_2) - t_{\alpha, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\mu_1 - \mu_2 \geq (7.8 - 5.6) - 1.734(2.55)\sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$\mu_1 - \mu_2 \geq 0.22$$

Because zero is not contained in this interval, we reject the null hypothesis.

c) $\Delta = 3$ Use s_p as an estimate of σ:

$$d = \frac{\mu_2 - \mu_1}{2s_p} = \frac{3}{2(2.55)} = 0.59$$

Using Chart VII (g) with d = 0.59 and $n = n_1 = n_2 = 10$ we obtain $n^* = 2n - 1 = 19$ and $\alpha = 0.05$. Therefore, $\beta \approx 0.18$ and the power is $1-\beta = 0.82$

d) $\beta = 0.05$, $d = \dfrac{2.5}{2(2.55)} = 0.49$, therefore $n^* \cong 45$. Finally, $n = \dfrac{n^*+1}{2} \cong 23$, and $n = n_1 = n_2 = 23$

10-19 a)

1) The parameter of interest is the difference in mean rod diameter, $\mu_1 - \mu_2$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,30} = -2.042$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,30} = 2.042$ for $\alpha = 0.05$

6) $\bar{x}_1 = 8.73$ $\bar{x}_2 = 8.68$

$s_1^2 = 0.35$ $s_2^2 = 0.40$

$n_1 = 15$ $n_2 = 17$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{14(0.35) + 16(0.40)}{30}} = 0.614$$

$$t_0 = \frac{(8.73 - 8.68)}{0.614\sqrt{\frac{1}{15} + \frac{1}{17}}} = 0.230$$

7) Conclusion: Because $-2.042 < 0.230 < 2.042$, fail to reject the null hypothesis. There is insufficient evidence to conclude that the two machines produce different mean diameters at $\alpha = 0.05$.

P-value = $2P(t > 0.230) > 2(0.40)$, P-value > 0.80

b) 95% confidence interval: $t_{0.025,30} = 2.042$

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(8.73-8.68) - 2.042(0.614)\sqrt{\frac{1}{15}+\frac{1}{17}} \le \mu_1 - \mu_2 \le \left(8.73-8.68\right) + 2.042(0.643)\sqrt{\frac{1}{15}+\frac{1}{17}}$$

$$-0.394 \le \mu_1 - \mu_2 \le 0.494$$

Because zero is contained in this interval, there is insufficient evidence to conclude that the two machines produce rods with different mean diameters.

10-20

a) Assume the populations follow normal distributions and $\sigma_1^2 = \sigma_2^2$. The assumption of equal variances may be relaxed in this case because it is known that the t-test and confidence intervals involving the t-distribution are robust to the assumption of equal variances when sample sizes are equal.

Case 1: AFFF	Case 2: ATC
μ_1 = mean foam expansion for AFFF	μ_2 = mean foam expansion for ATC
$\bar{x}_1 = 4.7$	$\bar{x}_2 = 6.9$
$s_1 = 0.6$	$s_2 = 0.8$
$n_1 = 5$	$n_2 = 5$

95% confidence interval: $t_{0.025,8} = 2.306$ $\quad s_p = \sqrt{\frac{4(0.60^2)+4(0.80^2)}{8}} = 0.7071$

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(4.7-6.9) - 2.306(0.7071)\sqrt{\frac{1}{5}+\frac{1}{5}} \le \mu_1 - \mu_2 \le \left(4.7-6.9\right) + 2.306(0.7071)\sqrt{\frac{1}{5}+\frac{1}{5}}$$

$$-3.23 \le \mu_1 - \mu_2 \le -1.17$$

b) Yes, with 95% confidence, the mean foam expansion for ATC exceeds that of AFFF by between 1.17 and 3.23 units.

10-21 a) 1) The parameter of interest is the difference in mean catalyst yield, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha,n_1+n_2-2}$ where $-t_{0.01,25} = -2.485$ for $\alpha = 0.01$

6) $\bar{x}_1 = 86$ $\quad \bar{x}_2 = 89$

$s_1 = 3$ $\quad s_2 = 2$

$n_1 = 12$ $\quad n_2 = 15$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{11(3)^2+14(2)^2}{25}} = 2.4899$$

$$t_0 = \frac{(86-89)}{2.4899\sqrt{\frac{1}{12}+\frac{1}{15}}} = -3.11$$

7) Conclusion: Because –3.11 < –2.485, reject the null hypothesis and conclude that the mean yield of catalyst 2 exceeds that of catalyst 1 at $\alpha = 0.01$.

b) 99% upper confidence interval $\mu_1 - \mu_2$: $t_{0.01,25} = 2.485$

$$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

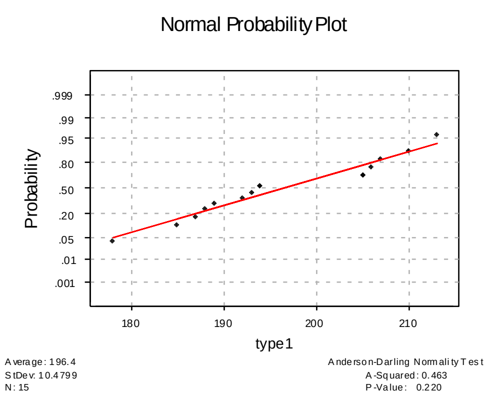
$$\mu_1 - \mu_2 \le (86 - 89) + 2.485(2.4899)\sqrt{\frac{1}{12} + \frac{1}{15}}$$

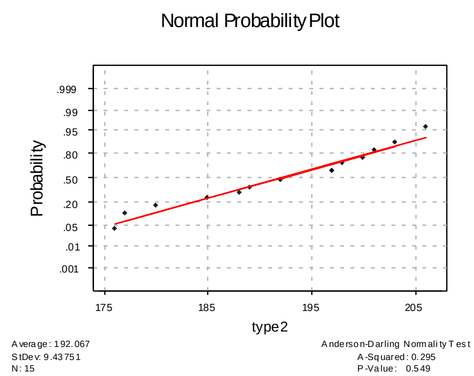
$$\mu_1 - \mu_2 \le -0.603 \text{ or equivalently } \mu_1 + 0.603 \le \mu_2$$

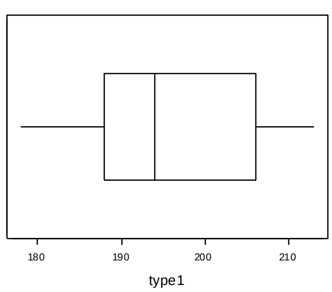
We are 99% confident that the mean yield of catalyst 2 exceeds that of catalyst 1 by at least 0.603 units.

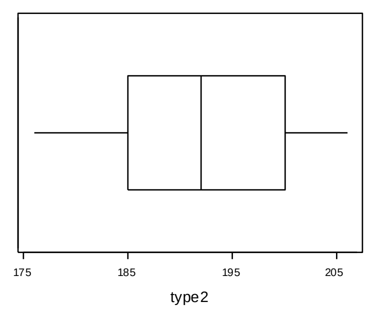
10-22

a) According to the normal probability plots, the assumption of normality is reasonable because the data fall approximately along lines. The equality of variances does not appear to be severely violated either because the slopes are approximately the same for both samples.









b) 1) The parameter of interest is the difference in deflection temperature under load, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 > 0$ or $\mu_1 > \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 > t_{\alpha, n_1+n_2-2}$ where $t_{0.05,28} = 1.701$ for $\alpha = 0.05$

6) <u>Type 1</u> <u>Type 2</u>

$\bar{x}_1 = 196.4$ $\bar{x}_2 = 192.067$

$s_1 = 10.48$ $s_2 = 9.44$

$n_1 = 15$ $n_2 = 15$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{14(10.48)^2 + 14(9.44)^2}{28}} = 9.97$$

$$t_0 = \frac{(196.4 - 192.067)}{9.97\sqrt{\frac{1}{15}+\frac{1}{15}}} = 1.19$$

7) Conclusion: Because 1.19 < 1.701 fail to reject the null hypothesis. There is insufficient evidence to conclude that the mean deflection temperature under load for type 1 exceeds the mean for type 2 at the 0.05 level of significance.

P-value = $P(t > 1.19)$, $0.1 < P\text{-value} < 0.25$

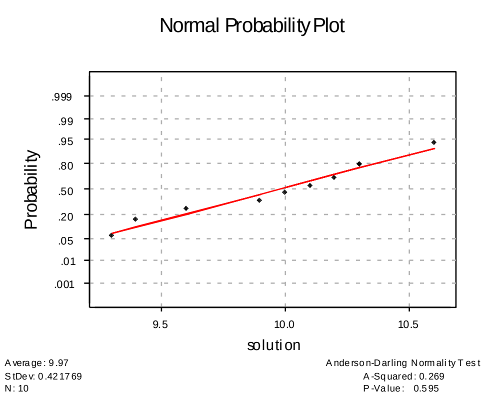
c) $\Delta = 5$ Use s_p as an estimate of σ:

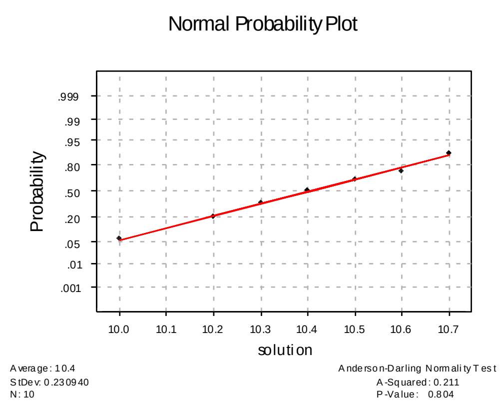
$$d = \frac{\mu_1 - \mu_2}{2s_p} = \frac{5}{2(9.97)} = 0.251$$

Using Chart VII (g) with $\beta = 0.10$, d = 0.251 we get $n^* \cong 100$. Because $n^* = 2n - 1$, $n_1 = n_2 = 50$. Therefore, the sample sizes of 15 are not adequate to meet the given probability of detection.

10-23

a) According to the normal probability plots, the assumption of normality appears to reasonable because the data from both the samples fall approximately along a line. The equality of variances does not appear to be severely violated either because the slopes are approximately the same for both samples.





b) 1) The parameter of interest is the difference in mean etch rate, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2,n_1+n_2-2}$ where $-t_{0.025,18} = -2.101$ or $t_0 > t_{\alpha/2,n_1+n_2-2}$ where $t_{0.025,18} = 2.101$ for $\alpha = 0.05$

6) $\bar{x}_1 = 9.97$ $\bar{x}_2 = 10.4$

$s_1 = 0.422$ $s_2 = 0.231$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{9(0.422)^2 + 9(0.231)^2}{18}} = 0.340$$

$n_1 = 10 \qquad n_2 = 10$

$$t_0 = \frac{(9.97 - 10.4)}{0.340\sqrt{\frac{1}{10} + \frac{1}{10}}} = -2.83$$

7) Conclusion: Because –2.83 < –2.101 reject the null hypothesis and conclude the two machines mean etch rates differ at $\alpha = 0.05$.

P-value = $2P(t < -2.83)$ 2(0.005) < *P*-value < 2(0.010) = 0.010 < *P*-value < 0.020

c) 95% confidence interval: $t_{0.025,18} = 2.101$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(9.97 - 10.4) - 2.101(.340)\sqrt{\frac{1}{10} + \frac{1}{10}} \le \mu_1 - \mu_2 \le (9.97 - 10.4) + 2.101(.340)\sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$-0.7495 \le \mu_1 - \mu_2 \le -0.1105$$

We are 95% confident that the mean etch rate for solution 2 exceeds the mean etch rate for solution 1 by between 0.1105 and 0.7495.

10-24 a) 1) The parameter of interest is the difference in mean impact strength, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha,\nu}$ where $t_{0.05,23} = 1.714$ for $\alpha = 0.05$ since

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)}{n_2 - 1}} = 23.72$$

$$\nu \cong 23$$

(truncated)

6) $\bar{x}_1 = 290 \qquad \bar{x}_2 = 321$

$s_1 = 12 \qquad s_2 = 22$

$n_1 = 10 \qquad n_2 = 16$

$$t_0 = \frac{(290 - 321)}{\sqrt{\frac{(12)^2}{10} + \frac{(22)^2}{16}}} = -4.64$$

7) Conclusion: Because –4.64 < –1.714 reject the null hypothesis and conclude that supplier 2 provides gears with higher mean impact strength at the 0.05 level of significance.

P-value = P(t < –4.64): *P*-value < 0.0005

b) 1) The parameter of interest is the difference in mean impact strength, $\mu_2 - \mu_1$

2) $H_0: \mu_2 - \mu_1 = 25$

3) H1 : $\mu_2 - \mu_1 > 25$ or $\mu_2 > \mu_1 + 25$
4) The test statistic is

$$t_0 = \frac{(\bar{x}_2 - \bar{x}_1) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 > t_{\alpha,\nu} = 1.714$ for $\alpha = 0.05$ where

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)}{n_2 - 1}} = 23.72$$

$$\nu \cong 23$$

6) $\bar{x}_1 = 290$ $\bar{x}_2 = 321$ $\Delta_0 = 25$ $s_1 = 12$ $s_2 = 22$ $n_1 = 10$ $n_2 = 16$

$$t_0 = \frac{(321 - 290) - 25}{\sqrt{\frac{(12)^2}{10} + \frac{(22)^2}{16}}} = 0.898$$

7) Conclusion: Because 0.898 < 1.714, fail to reject the null hypothesis. There is insufficient evidence to conclude that t the mean impact strength from supplier 2 is at least 25 ft-lb higher that supplier 1 using $\alpha = 0.05$.

c) Using the information provided in part (a), and $t_{0.025,25} = 2.069$, a 95% confidence interval on the difference $\mu_2 - \mu_1$ is

$$(\bar{x}_2 - \bar{x}_1) - t_{0.025,25}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_2 - \mu_1 \le (\bar{x}_2 - \bar{x}_1) + t_{0.025,25}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$31 - 2.069(6.682) \le \mu_2 - \mu_1 \le 31 + 2.069(6.682)$$

$$17.175 \le \mu_2 - \mu_1 \le 44.825$$

Because zero is not contained in the confidence interval, we conclude that supplier 2 provides gears with a higher mean impact strength than supplier 1 with 95% confidence.

10-25 a)
1) The parameter of interest is the difference in mean melting point, $\mu_1 - \mu_2$, with $\Delta_0 = 0$
2) H0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) H1 : $\mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$
4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2,n_1+n_2-2}$ where $-t_{0.0025,40} = -2.021$ or $t_0 > t_{\alpha/2,n_1+n_2-2}$ where $t_{0.025,40} = 2.021$ for $\alpha = 0.05$

6) $\bar{x}_1 = 420$ $\bar{x}_2 = 426$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$$

$s_1 = 4$ $s_2 = 3$

$$= \sqrt{\frac{20(4)^2 + 20(3)^2}{40}} = 3.536$$

$n_1 = 21$ $n_2 = 21$

$$t_0 = \frac{(420 - 426)}{3.536\sqrt{\frac{1}{21} + \frac{1}{21}}} = -5.498$$

7) Conclusion: Because –5.498 < –2.021 reject the null hypothesis. The alloys differ significantly in mean melting point at α = 0.05.

P-value = $2P(t < -5.498)$ *P*-value < 0.0010

b) $d = \dfrac{|\mu_1 - \mu_2|}{2\sigma} = \dfrac{3}{2(4)} = 0.375$

Using the appropriate chart in the Appendix, with β = 0.10 and α = 0.05 we have n^* = 75.

Therefore, $n = \dfrac{n^* + 1}{2} = 38$, $n_1 = n_2 = 38$

10-26 a)

1) The parameter of interest is the difference in mean speed, $\mu_1 - \mu_2$, $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 > 0$ or $\mu_1 > \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 > t_{\alpha, n_1+n_2-2}$ where $t_{0.10,14} = 1.345$ for α = 0.10

6) <u>Case 1: 25 mil</u> <u>Case 2: 20 mil</u>

$\bar{x}_1 = 1.15$ $\bar{x}_2 = 1.06$ $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

$s_1 = 0.11$ $s_2 = 0.09$ $= \sqrt{\dfrac{7(0.11)^2 + 7(0.09)^2}{14}} = 0.1005$

$n_1 = 8$ $n_2 = 8$

$$t_0 = \frac{(1.15 - 1.06)}{0.1005\sqrt{\frac{1}{8} + \frac{1}{8}}} = 1.79$$

7) Because 1.79 > 1.345 reject the null hypothesis and conclude that reducing the film thickness from 25 mils to 20 mils significantly increases the mean speed of the film at the 0.10 level of significance (Note: an increase in film speed will result in *lower* values of observations).

P-value = $P(t > 1.79)$ 0.025 < *P*-value < 0.05

b) 95% confidence interval: $t_{0.025,14} = 2.145$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(1.15 - 1.06) - 2.145(0.1005)\sqrt{\frac{1}{8} + \frac{1}{8}} \le \mu_1 - \mu_2 \le (1.15 - 1.06) + 2.145(0.1005)\sqrt{\frac{1}{8} + \frac{1}{8}}$$

$$-0.0178 \le \mu_1 - \mu_2 \le 0.1978$$

We are 90% confident the difference in mean speed of the film is between -0.0178 and 0.1978 $\mu J/in^2$.

10-27 a)

1) The parameter of interest is the difference in mean wear amount, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

. 2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{0.025,26}$ or $t_0 > t_{0.025,26}$ where $t_{0.025,26} = 2.056$ for $\alpha = 0.05$ because

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)}{n_2 - 1}} = 26.98$$

$$\nu \cong 26$$

6) $\bar{x}_1 = 20$ $\bar{x}_2 = 15$

$s_1 = 2$ $s_2 = 8$

$n_1 = 25$ $n_2 = 25$

$$t_0 = \frac{(20 - 15)}{\sqrt{\frac{(2)^2}{25} + \frac{(8)^2}{25}}} = 3.03$$

7) Conclusion: Because 3.03 > 2.056 reject the null hypothesis. The data support the claim that the two companies produce material with significantly different wear at the 0.05 level of significance.

P-value = 2P(t > 3.03), 2(0.0025) < *P*-value < 2(0.005), 0.005 < *P*-value < 0.010

b)
1) The parameter of interest is the difference in mean wear amount, $\mu_1 - \mu_2$

2) $H_0: \mu_1 - \mu_2 = 0$

3) $H_1: \mu_1 - \mu_2 > 0$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 > t_{0.05,27}$ where $t_{0.05,26} = 1.706$ for $\alpha = 0.05$ since

6) $\bar{x}_1 = 20$ $\bar{x}_2 = 15$

$s_1 = 2$ $s_2 = 8$

$n_1 = 25$ $n_2 = 25$

$$t_0 = \frac{(20 - 15)}{\sqrt{\frac{(2)^2}{25} + \frac{(8)^2}{25}}} = 3.03$$

7) Conclusion: Because 3.03 > 1.706 reject the null hypothesis. The data support the claim that the material from company 1 has a higher mean wear than the material from company 2 at a 0.05 level of significance.

c) For part (a) use a 95% two-sided confidence interval:
$t_{0.025,26} = 2.056$

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(20 - 15) - 2.056\sqrt{\frac{(2)^2}{25} + \frac{(8)^2}{25}} \le \mu_1 - \mu_2 \le (20 - 15) + 2.056\sqrt{\frac{(2)^2}{25} + \frac{(8)^2}{25}}$$

$$1.609 \le \mu_1 - \mu_2 \le 8.391$$

For part (b) use a 95% lower one-sided confidence interval:

$t_{0.05,26} = 1.706$

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2$$

$$(20 - 15) - 1.706\sqrt{\frac{(2)^2}{25} + \frac{(8)^2}{25}} \le \mu_1 - \mu_2$$

$$2.186 \le \mu_1 - \mu_2$$

For part a) we are 95% confident the mean abrasive wear from company 1 exceeds the mean abrasive wear from company 2 by between 1.609 and 8.391 mg/1000.

For part b) we are 95% confident the mean abrasive wear from company 1 exceeds the mean abrasive wear from company 2 by at least 2.186 mg/1000.

10-28 a)

1) The parameter of interest is the difference in mean coating thickness, $\mu_1 - \mu_2$, with $\Delta_0 = 0$.

2) $H_0: \mu_1 - \mu_2 = 0$

3) $H_1: \mu_1 - \mu_2 > 0$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 > t_{0.01,18}$ where $t_{0.01,18} = 2.552$ for $\alpha = 0.01$ since

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)}{n_2 - 1}} 18.37$$

$$\nu \cong 18$$

(truncated)

6) $\bar{x}_1 = 103.5$ $\bar{x}_2 = 99.7$

$s_1 = 10.2$ $s_2 = 20.1$

$n_1 = 11$ $n_2 = 13$

$$t_0 = \frac{(103.5 - 99.7)}{\sqrt{\frac{(10.2)^2}{11} + \frac{(20.1)^2}{13}}} = 0.597$$

7) Conclusion: Because 0.597 < 2.552, fail to reject the null hypothesis. There is insufficient evidence to conclude that increasing the temperature reduces the mean coating thickness at $\alpha = 0.01$.

P-value = P(t > 0.597), 0.25 < *P*-value < 0.40

b) If $\alpha = 0.01$, construct a 99% two-sided confidence interval on the difference in means.

Here, $t_{0.005,19}$ = 2.878

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(103.5 - 99.7) - 2.878\sqrt{\frac{(10.2)^2}{11} + \frac{(20.1)^2}{13}} \le \mu_1 - \mu_2 \le (103.5 - 99.7) - 2.878\sqrt{\frac{(10.2)^2}{11} + \frac{(20.1)^2}{13}}$$

$$-14.52 \le \mu_1 - \mu_2 \le 22.12$$

Because the interval contains zero, there is not a significant difference in the mean coating thickness between the two temperatures.

10-29

a)

1) The parameter of interest is the difference in mean width of the backside chip-outs for the single spindle saw process versus the dual spindle saw process , $\mu_1 - \mu_2$

2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 \ne 0$ or $\mu_1 \ne \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2,n_1+n_2-2}$ where $-t_{0.025,28}$ = –2.048 or $t_0 > t_{\alpha/2,n_1+n_2-2}$ where $t_{0.025,28}$ = 2.048 for α = 0.05

6) $\bar{x}_1 = 66.385$ $\bar{x}_2 = 45.278$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$s_1^2 = 7.895^2$ $s_2^2 = 8.612^2$

$$= \sqrt{\frac{14(7.895)^2 + 14(8.612)^2}{28}} = 8.26$$

n_1 = 15 n_2 = 15

$$t_0 = \frac{(66.385 - 45.278)}{8.26\sqrt{\dfrac{1}{15} + \dfrac{1}{15}}} = 7.00$$

7) Conclusion: Because 7.00 > 2.048, we reject the null hypothesis at α = 0.05. *P*-value ≅ 0

b) 95% confidence interval: $t_{0.025,28}$ = 2.048

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(66.385 - 45.278) - 2.048(8.26)\sqrt{\frac{1}{15} + \frac{1}{15}} \le \mu_1 - \mu_2 \le (66.385 - 45.278) + 2.048(8.26)\sqrt{\frac{1}{15} + \frac{1}{15}}$$

$$14.93 \le \mu_1 - \mu_2 \le 27.28$$

Because zero is not contained in this interval, we reject the null hypothesis.

c) For β < 0.01 and d = $\dfrac{15}{2(8.26)}$ = 0.91, with α = 0.05 then using Chart VII (e) we find n^* > 15. Then $n > \dfrac{15+1}{2} = 8$

10-30 a)

1) The parameter of interest is the difference in mean blood pressure between the test and control groups, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha,\nu}$ where $t_{0.05,12}$ = -1.782 for α = 0.05 since

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)}{n_2 - 1}} = 12$$

$$\nu \cong 12$$

6)

$\bar{x}_1 = 90$ $\bar{x}_2 = 115$

$s_1 = 5$ $s_2 = 10$

$n_1 = 8$ $n_2 = 9$

$$t_0 = \frac{(90 - 115)}{\sqrt{\frac{(5)^2}{8} + \frac{(10)^2}{9}}} = -6.63$$

7) Conclusion: Because –6.62 < –1.782 reject the null hypothesis and conclude that the test group has higher mean arterial blood pressure than the control group at the 0.05 level of significance.

P-value = P(t < –6.62): *P*-value $\cong$ 0

b) 95% confidence interval: $t_{0.05,12}$ = 1.782

$$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\mu_1 - \mu_2 \le (90 - 115) + 1.782\sqrt{\frac{5^2}{8} + \frac{10^2}{9}}$$

$$\mu_1 - \mu_2 \le -18.28$$

Because zero is not contained in this interval, we reject the null hypothesis.

c)

1) The parameter of interest is the difference in mean blood pressure between the test and control groups, $\mu_1 - \mu_2$, with $\Delta_0 = -15$

2) H_0 : $\mu_1 - \mu_2 = -15$ or $\mu_1 = \mu_2$

3) H_1 : $\mu_1 - \mu_2 < -15$ or $\mu_1 < \mu_2 - 15$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha,\nu}$ where $t_{0.05,12}$ = -1.782 for α = 0.05 since

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)}{n_2 - 1}} = 12$$

$$\nu \cong 12$$

6) $\bar{x}_1 = 90$ $\bar{x}_2 = 115$

$s_1 = 5$ $s_2 = 10$

$n_1 = 8$ $n_2 = 9$

$$t_0 = \frac{(90 - 115) + 15}{\sqrt{\frac{(5)^2}{8} + \frac{(10)^2}{9}}} = -2.65$$

7) Conclusion: Because –2.65 < –1.782 reject the null hypothesis and conclude that the test group has higher mean arterial blood pressure than the control group at the 0.05 level of significance.

d) 95% confidence interval: $t_{0.05,12}$ = 1.782

$$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\mu_1 - \mu_2 \le -18.28$$

Because -15 is greater than the values in this interval, we are 95% confident that the mean for the test group is at least 15 mmHg higher than the control group.

10-31 a)

1) The parameter of interest is the difference in mean number of periods in a sample of 200 trains for two different levels of noise voltage, 100mv and 150mv

$\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 > 0$ or $\mu_1 > \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 > t_{\alpha, n_1+n_2-2}$ where $t_{0.01,198}$ = 2.326 for α = 0.01

6)

$\bar{x}_1 = 7.9$ $\bar{x}_2 = 6.9$ $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

$s_1 = 2.6 \qquad s_2 = 2.4$

$n_1 = 100 \qquad n_2 = 100$

$$= \sqrt{\frac{99(2.6)^2 + 99(2.4)^2}{198}} = 2.5$$

$$t_0 = \frac{(7.9 - 6.9)}{2.5\sqrt{\frac{1}{100} + \frac{1}{100}}} = 2.82$$

7) Conclusion: Because 2.82 > 2.326, reject the null hypothesis at the 0.01 level of significance.

P-value = $P(t > 2.82)$ *P*-value ≅ 0.0025

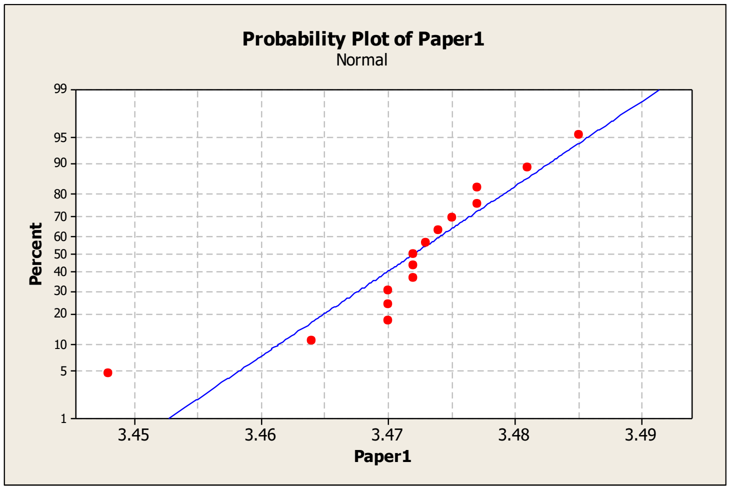
b) 99% confidence interval: $t_{0.01,198} = 2.326$

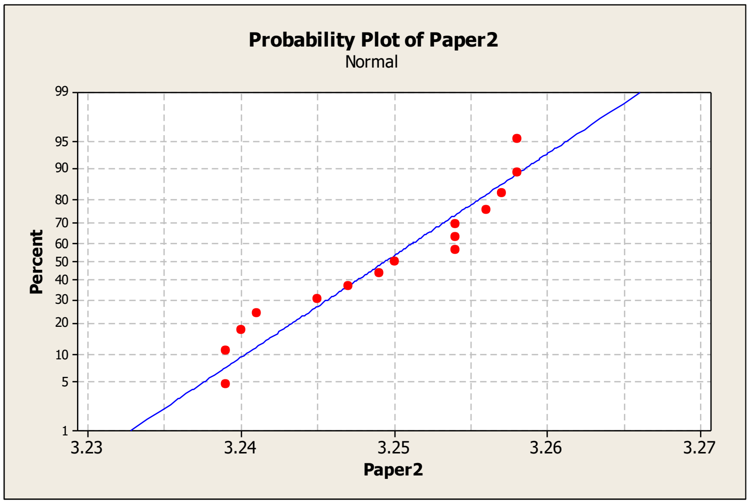
$$\mu_1 - \mu_2 \geq (\bar{x}_1 - \bar{x}_2) - t_{\alpha, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\mu_1 - \mu_2 \geq 0.178$$

Because zero is not contained in this interval, reject the null hypothesis.

10-32 a) The probability plots below show that the normality assumptions are reasonable for both data sets.





b)
1) The parameter of interest is the difference in mean weight of two sheets of paper, $\mu_1 - \mu_2$. Assume equal variances.
2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$
4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,28} = -2.048$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where

$t_{0.025,28} = 2.048$ for $\alpha = 0.05$

6) $\bar{x}_1 = 3.472$ $\bar{x}_2 = 3.2494$

$s_1^2 = 0.00831^2$ $s_2^2 = .00714^2$

$n_1 = 15$ $n_2 = 15$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{14(.00831)^2 + 14(0.00714)^2}{28}} = .00775$$

$t_0 = 78.66$

7) Conclusion: Because 78.66 >2.048, reject the null hypothesis at $\alpha = 0.05$.

P-value $\cong 0$

c)
1) The parameter of interest is the difference in mean weight of two sheets of paper, $\mu_1 - \mu_2$
2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$
4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.05,28} = -1.701$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.05,28}$ = 1.701 for α = 0.1

6) $\bar{x}_1 = 3.472$ $\bar{x}_2 = 3.2494$

$s_1^2 = 0.00831^2$ $s_2^2 = .00714^2$

n_1 = 15 n_2 = 15

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{14(.00831)^2 + 14(0.00714)^2}{28}} = .00775$$

$t_0 = 78.66$

7) Conclusion: Because 78.66 > 1.701, reject the null hypothesis at α = 0.10. *P*-value ≅ 0

d) The answer is the same because the decision to reject the null hypothesis made in part (b) was at a lower level of significance than the test in (c). Therefore, the decision is the same for any value of α larger than that used in part (b). Alternatively, the P-value from part (b) is essentially 0, meaning that for any level of α greater than or equal to the P-value, the decision is to reject the null hypothesis.

e) 95% confidence interval for part (b): $t_{0.025,28} = 2.048$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$0.216 \le \mu_1 - \mu_2 \le 0.228$$

Because zero is not contained in this interval we reject the null hypothesis.

90% confidence interval for part (c): $t_{0.05,28} = 1.701$

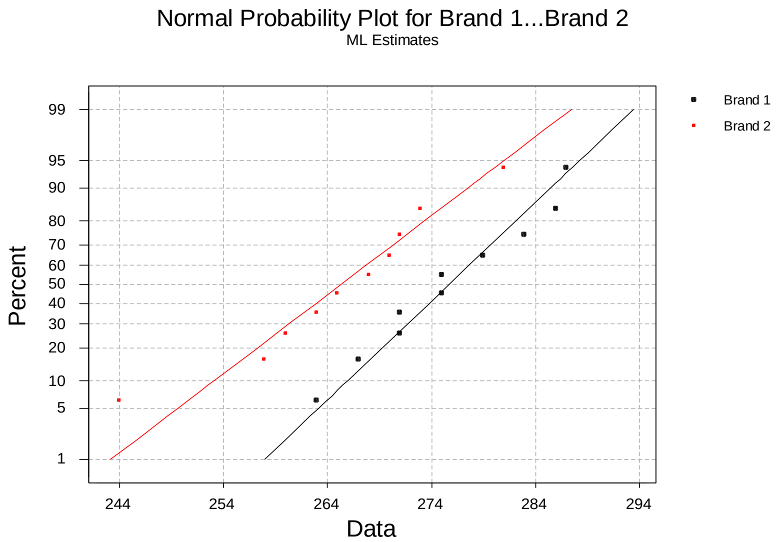
$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$0.217 \le \mu_1 - \mu_2 \le 0.227$$

Because zero is not contained in this interval we reject the null hypothesis.

10-33

a) The data appear to be normally distributed and the variances appear to be approximately equal. The slopes of the lines on the normal probability plots are almost the same.



b)

1) The parameter of interest is the difference in mean overall distance, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,18} = 2.101$ for $\alpha = 0.05$

6) $\bar{x}_1 = 275.7$ $\bar{x}_2 = 265.3$

$s_1 = 8.03$ $s_2 = 10.04$

$n_1 = 10$ $n_2 = 10$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{9(8.03)^2 + 9(10.04)^2}{20}} = 9.09$$

$$t_0 = \frac{(275.7 - 265.3)}{9.09\sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.558$$

7) Conclusion: Because 2.558 > 2.101, reject the null hypothesis. The data support the claim that the means differ at α = 0.05.

P-value = $2P(t > 2.558)$ *P*-value ≈ 2(0.01) = 0.02

c) $$(\bar{x}_1 - \bar{x}_2) - t_{\alpha,\nu} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha,\nu} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(275.7 - 265.3) - 2.101(9.09)\sqrt{\frac{1}{10} + \frac{1}{10}} \le \mu_1 - \mu_2 \le (275.7 - 265.3) + 2.101(9.09)\sqrt{\frac{1}{10} + \frac{1}{10}}$$

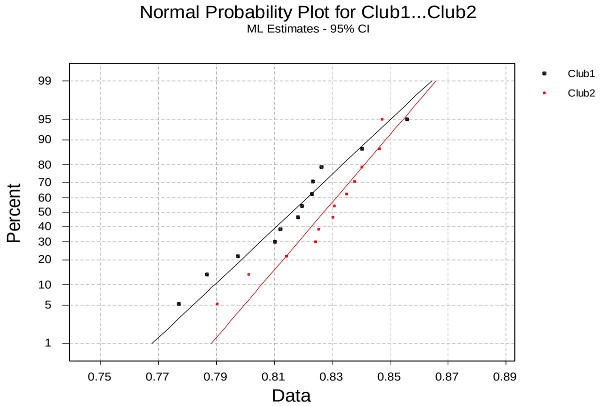
$$1.86 \le \mu_1 - \mu_2 \le 18.94$$

d) $d = \frac{5}{2(9.09)} 0.275$ β=0.95 Power = 1 - 0.95 = 0.05

e) $\beta = 0.25$ $d = \frac{3}{2(9.09)} = 0.165$ n*=100 Therefore, n = 51

10-34

a) The data appear to be normally distributed and the variances appear to be approximately equal. The slopes of the lines on the normal probability plots are almost the same.



b)

1) The parameter of interest is the difference in mean coefficient of restitution, $\mu_1 - \mu_2$

2) H₀ : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H₁ : $\mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,22} = 2.074$ for $\alpha = 0.05$

6) $\bar{x}_1 = 0.8161$ $\quad \bar{x}_2 = 0.8271$ $\quad s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

$s_1 = 0.0217$ $\quad s_2 = 0.0175$ $\quad = \sqrt{\frac{11(0.0217)^2 + 11(0.0175)^2}{22}} = 0.01971$

$n_1 = 12$ $\quad n_2 = 12$

$$t_0 = \frac{(0.8161 - 0.8271)}{0.01971\sqrt{\frac{1}{12} + \frac{1}{12}}} = -1.367$$

7) Conclusion: Because –1.367 > –2.074 fail to reject the null hypothesis. The data do not support the claim that there is a difference in the mean coefficients of restitution for club1 and club2 at $\alpha = 0.05$

P-value = 2P$\left(t < -1.36\right)$, *P*-value ≈ 2(0.1) = 0.2

c) 95% confidence interval

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha,\nu} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha,\nu} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(0.8161 - 0.8271) - 2.074(0.01971)\sqrt{\frac{1}{12} + \frac{1}{12}} \le \mu_1 - \mu_2 \le (0.8161 - 0.8271) + 2.074(0.01971)\sqrt{\frac{1}{12} + \frac{1}{12}}$$

$$-0.0277 \le \mu_1 - \mu_2 \le 0.0057$$

Because zero is included in the confidence interval there is not a significant difference in the mean coefficients of restitution at $\alpha = 0.05$.

d) $d = \frac{0.2}{2(0.01971)} = 5.07$ $\quad \beta \cong 0$, Power $\cong 1$

e) $1 - \beta = 0.8$ $\beta = 0.2$ $d = \dfrac{0.1}{2(0.01971)} = 2.53$ $n^* = 4$, $n = \dfrac{n^*+1}{2} = 2.5$ $n \cong 3$

10-35 The sample standard deviations differ substantially. Do not assume the populations standard deviations are equal.

$\bar{x}_1 = 190$ $\bar{x}_2 = 310$

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} = 3.22$$

$s_1 = 15$ $s_2 = 50$
$n_1 = 10$ $n_2 = 4$

$t_{0.05/2,\,3} = 3.182$

$$(\bar{x}_1 - \bar{x}_2) - t_{0.05/2,3}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{0.05/2,3}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(190 - 310) - 3.182 \times \sqrt{\frac{225}{10} + \frac{2500}{4}} \le \mu_1 - \mu_2 \le (190 - 310) + 3.182 \times \sqrt{\frac{225}{10} + \frac{2500}{4}}$$

$$-200.98 \le \mu_1 - \mu_2 \le -39.02$$

Because the confidence interval contains only negative values, the absorbency of towel 1 is less than that of towel 2. The confidence interval does not contain zero. Therefore, the hypothesis test for the equality of means would reject the null hypothesis at $\alpha = 0.05$.

10-36 a)

1) The parameter of interest is the difference in mean algae content, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \ne 0$ or $\mu_1 \ne \mu_2$. The alternative is two-sided.

4) The test statistic is (assume equal variances)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,26} = -2.056$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,26} = 2.056$ for $\alpha = 0.05$

6) $\bar{x}_1 = 44.58$ $\bar{x}_2 = 35.48$

$s_1 = 19.35$ $s_2 = 14.50$

$n_1 = 15$ $n_2 = 13$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{14(19.35)^2 + 12(14.50)^2}{26}} = 17.282$$

$$t_0 = \frac{(44.58 - 35.48)}{17.282\sqrt{\dfrac{1}{15} + \dfrac{1}{13}}} = 1.39$$

7) Conclusion: Because -2.056 < 1.39 < 2.056 fail to reject the null hypothesis. The rivers do not differ significantly in mean algae content at $\alpha = 0.05$.

b) Use a 95% two-sided confidence interval

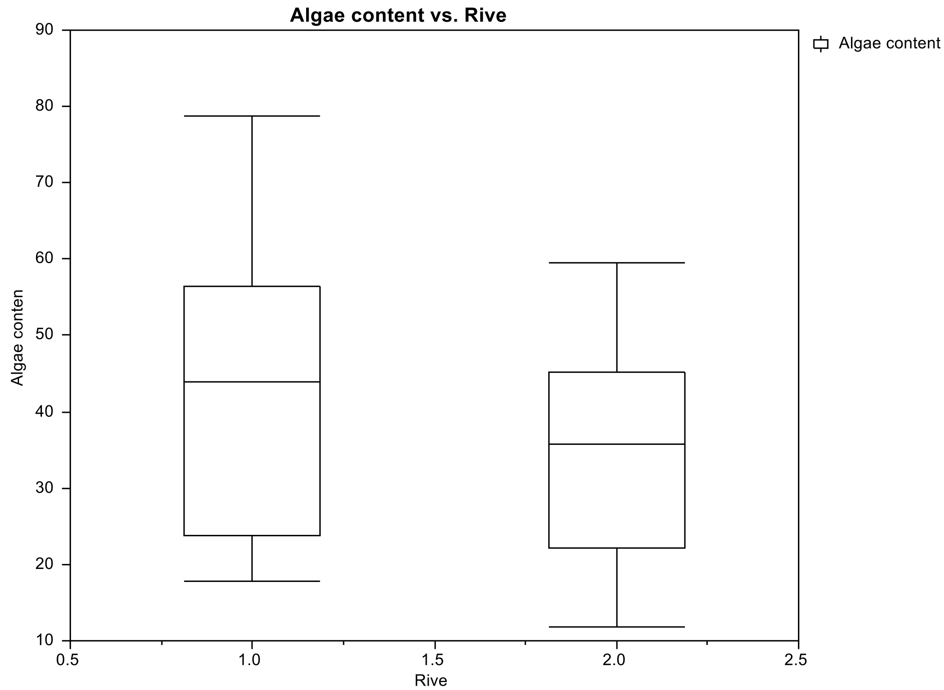
$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(44.58 - 35.48) - 2.056 \times 17.282 \times \sqrt{\frac{1}{15} + \frac{1}{13}} \le \mu_1 - \mu_2 \le (44.58 - 35.48) + 2.056 \times 17.282 \times \sqrt{\frac{1}{15} + \frac{1}{13}}$$

$-4.37 \le \mu_1 - \mu_2 \le 22.56$

c) Zero is contained in this interval. Therefore, we fail to reject the null hypothesis. There is no significant difference between the means.

d) We assumed that algae content of both rives have the same variance. However, river 2 has slightly greater sample variability.



10-37 Heats 5 and 6 are group 1 and heat 7 is group 2

1) The parameter of interest is the difference in time, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \ne 0$ or $\mu_1 \ne \mu_2$. The alternative is two-sided.

4) The test statistic is (assume equal variances)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,18} = -2.101$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,18} = 2.101$ for $\alpha = 0.05$

6) $\bar{x}_1 = 49.12$ $\bar{x}_2 = 48.91$

$s_1 = 0.63$ $s_2 = 0.24$

$n_1 = 13$ $n_2 = 7$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{12(0.63)^2 + 6(0.24)^2}{18}} = 0.535$$

$$t_0 = \frac{(49.12 - 48.91)}{0.535\sqrt{\frac{1}{13} + \frac{1}{7}}} = 0.85$$

7) Conclusion: Because -2.101 < 0.85 < 2.101 fail to reject the null hypothesis. The mean times do not differ significantly at $\alpha = 0.05$.

Heat 5 is group 1 and heat 7 is group 2

1) The parameter of interest is the difference in time, $\mu_1 - \mu_2$, with $\Delta_0 = 0$
2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$. The alternative is two-sided.
4) The test statistic is (assume equal variances)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,18} = -2.201$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,18} = 2.201$ for $\alpha = 0.05$

6) $\bar{x}_1 = 49.65$ $\quad \bar{x}_2 = 48.91$ $\quad s_1 = 0.38$ $\quad s_2 = 0.24$ $\quad n_1 = 6$ $\quad n_2 = 7$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{5(0.38)^2 + 6(0.24)^2}{11}} = 0.313$$

$$t_0 = \frac{(49.65 - 48.91)}{0.313\sqrt{\frac{1}{6} + \frac{1}{7}}} = 4.28$$

7) Conclusion: Because 2.101 < 4.28, reject the null hypothesis. The mean times differ significantly at $\alpha = 0.05$.

Heat 6 is group 1 and heat 7 is group 2
1) The parameter of interest is the difference in time, $\mu_1 - \mu_2$, with $\Delta_0 = 0$
2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$. The alternative is two-sided.
4) The test statistic is (assume equal variances)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,18} = -2.179$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,18} = 2.179$ for $\alpha = 0.05$

6) $\bar{x}_1 = 48.67$ $\quad \bar{x}_2 = 48.91$ $\quad s_1 = 0.40$ $\quad s_2 = 0.24$ $\quad n_1 = 7$ $\quad n_2 = 7$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{6(0.40)^2 + 6(0.24)^2}{12}} = 0.329$$

$$t_0 = \frac{(48.67 - 48.91)}{0.329\sqrt{\frac{1}{7} + \frac{1}{7}}} = -1.37$$

7) Conclusion: Because $-2.179 < -1.37 < 2.179$, fail to reject the null hypothesis. The mean times do not differ significantly at $\alpha = 0.05$.

10-38 a)
1) The parameter of interest is the difference in mean cycles to failure, $\mu_1 - \mu_2$, with $\Delta_0 = 0$
2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$. The alternative is two-sided.
4) The test statistic is (assume equal variances)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,18} = -2.101$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,18} = 2.101$ for $\alpha = 0.05$

6) $\bar{x}_1 = 422.20$ $\bar{x}_2 = 375.70$

$s_1 = 172.23$ $s_2 = 161.92$

$n_1 = 10$ $n_2 = 10$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{9(172.23)^2 + 9(161.92)^2}{18}} = 167.155$$

$$t_0 = \frac{(422.2 - 375.7)}{167.155\sqrt{\frac{1}{10} + \frac{1}{10}}} = 0.622$$

7) Conclusion: Because -2.101 < 0.622 < 2.101 fail to reject the null hypothesis. The temperatures do not differ significantly in mean cycles to failure at $\alpha = 0.05$.

b) Use a 95% two-sided confidence interval

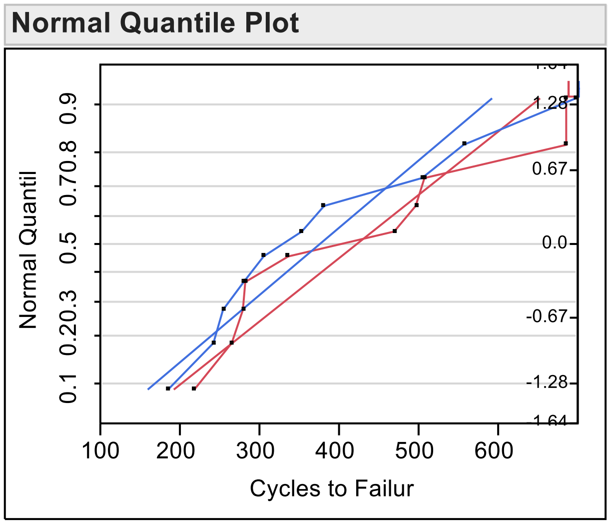
$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(422.2 - 375.7) - 2.101 \times 167.155 \times \sqrt{\frac{1}{10} + \frac{1}{10}} \le \mu_1 - \mu_2 \le (422.2 - 375.7) + 2.101 \times 167.155 \times \sqrt{\frac{1}{10} + \frac{1}{10}}$$

$$-110.56 \le \mu_1 - \mu_2 \le 203.56$$

c) Zero is contained in this interval. As a result, we conclude that there is no significant difference between the means and we fail to reject the null hypothesis.

d) Normal probability plots are shown below (blue line is for 60° degrees). According to these plots, the assumption of normality is reasonable because the data fall approximately along lines. The equality of variances does not appear to be severely violated either because the slopes are approximately the same for both samples.



10-39 a)

1) The parameter of interest is the difference in mean compression between storage conditions, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) H1 : $\mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is (assume equal variance)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha,n_1+n_2-2}$ where $t_{0.05,16}$ =1.746 for α = 0.05

6) <u>Case 1: 50°C</u> <u>Case 2: 60°C</u>

$\bar{x}_1 = 0.096$ $\bar{x}_2 = 0.179$

$s_1 = 0.049$ $s_2 = 0.094$

$n_1 = 9$ $n_2 = 9$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{8(0.049)^2 + 8(0.094)^2}{16}} = 0.075$$

$$t_0 = \frac{(0.096 - 0.179)}{0.075\sqrt{\frac{1}{9} + \frac{1}{9}}} = -2.349$$

7) Conclusion: Because –2.349 < –1.746, reject the null hypothesis and conclude that the mean compression increases with temperature at a 0.05 level of significance.

b) 95% one-sided confidence interval: $t_{0.05,16}$ = 1.746

$$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\mu_1 - \mu_2 \le (0.096 - 0.179) + 1.746(0.075)\sqrt{\frac{1}{9} + \frac{1}{9}}$$

$$\mu_1 - \mu_2 \le -0.021$$

95% two-sided confidence interval

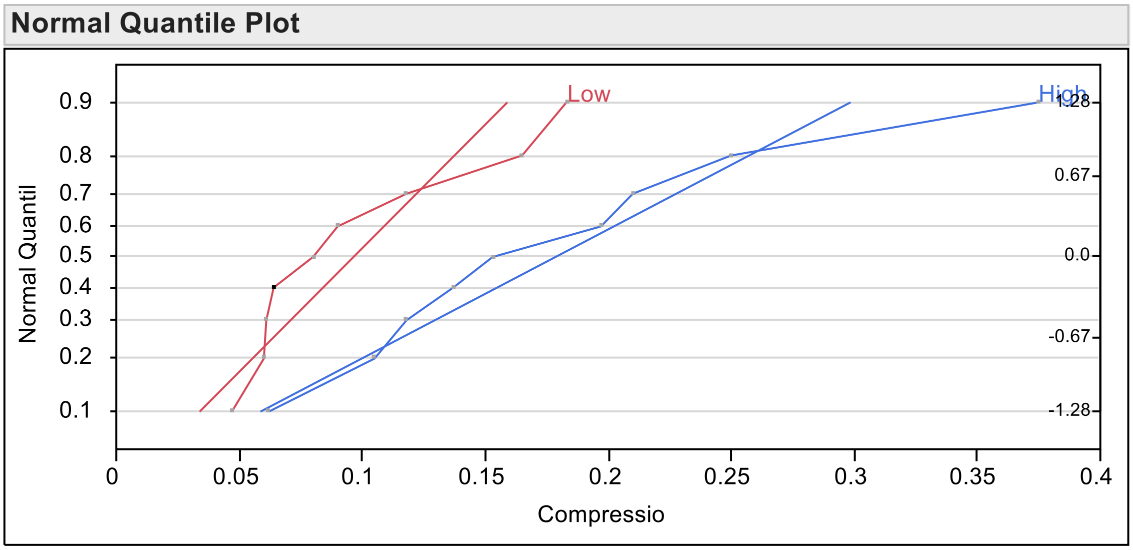
$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(0.096 - 0.179) - 2.12(0.075)\sqrt{\frac{1}{9} + \frac{1}{9}} \le \mu_1 - \mu_2 \le (0.096 - 0.179) + 2.12(0.075)\sqrt{\frac{1}{9} + \frac{1}{9}}$$

$$-0.158 \le \mu_1 - \mu_2 \le -0.008$$

c) Because the test is one-sided, we consider the one-sided confidence interval. Because zero is not contained in this interval, we reject the null hypothesis.

d) According to the normal probability plots, the assumption of normality is reasonable because the data fall approximately along lines. However, the slopes do not appear to be equal. Therefore, it is not very reasonable to assume that $\sigma_1^2 = \sigma_2^2$.



10-40 a)

1) The parameter of interest is the difference in mean grinding force between vibration levels, $\mu_1 - \mu_2$, with $\Delta_0 = 0$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 < 0$ or $\mu_1 < \mu_2$

4) The test statistic is (assume equal variance)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha, n_1+n_2-2}$ where $t_{0.05,22}$ =1.717 for $\alpha = 0.05$

6) Case 1: Low Case 2: High

$\bar{x}_1 = 258.917$ $\bar{x}_2 = 353.667$ $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

$s_1 = 26.221$ $s_2 = 46.596$ $= \sqrt{\frac{12(26.221)^2 + 12(46.596)^2}{22}} = 37.807$

$n_1 = 12$ $n_2 = 12$

$$t_0 = \frac{(258.917 - 353.667)}{37.807\sqrt{\frac{1}{12} + \frac{1}{12}}} = -6.139$$

7) Conclusion: Because $-6.139 < -1.717$ reject the null hypothesis and conclude that the mean grinding force increases with vibration level at a 0.05 level of significance.

b) 95% one-sided confidence interval: $t_{0.05,22} = 1.717$

$$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha, n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\mu_1 - \mu_2 \le (258.917 - 353.667) + 1.717(37.807)\sqrt{\frac{1}{12} + \frac{1}{12}}$$

$$\mu_1 - \mu_2 \le -68.249$$

95% two-sided confidence interval

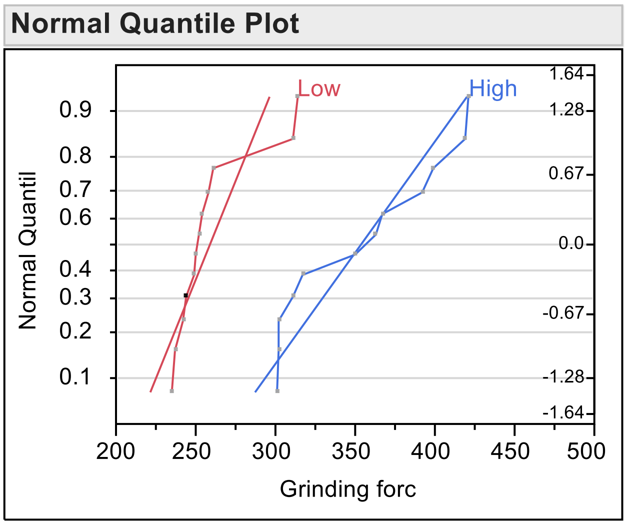
$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le\mu_1-\mu_2\le(\bar{x}_1-\bar{x}_2)+t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(258.917-353.667)-2.074(37.807)\sqrt{\frac{1}{12}+\frac{1}{12}}\le\mu_1-\mu_2\le(258.917-353.667)+2.074(37.807)\sqrt{\frac{1}{12}+\frac{1}{12}}$$

$$-126.759\le\mu_1-\mu_2\le-62.741$$

c) Because the test is one-sided, we consider the one-sided confidence interval. Because zero is not contained in this interval, we reject the null hypothesis.

d) According to the normal probability plots, the assumption of normality is reasonable because the data fall approximately along lines. The slopes appear to be similar, so it is reasonable to assume that $\sigma_1^2=\sigma_2^2$.



Section 10-3

10-41 a)

1) The parameters of interest are the mean current (note: set circuit 1 equal to sample 2 so that Table X can be used. Therefore, μ_1 = mean of circuit 2 and μ_2 = mean of circuit 1)

2) $H_0: \mu_1=\mu_2$

3) $H_1: \mu_1>\mu_2$

4) The test statistic is $w_2=\frac{(n_1+n_2)(n_1+n_2+1)}{2}-w_1$

5) Reject H_0 if $w_2 \le w^*_{0.025}$ = 51. Because α = 0.025 and n_1 = 8 and n_2 = 9, Appendix A, Table X gives the critical value.

6) w_1 = 78 and w_2 = 75 and because 75 is less than, 51 fail to reject H_0

7) Conclusion, fail to reject H_0. There is not enough evidence to conclude that the mean of circuit 2 exceeds the mean of circuit 1.

b)

1) The parameters of interest are the mean image brightness of the two tubes

2. $H_0: \mu_1=\mu_2$

3. $H_1: \mu_1>\mu_2$

4) The test statistic is $z_0=\frac{W_1-\mu_{w_1}}{\sigma_{w_1}}$

5) We reject H_0 if $Z_0 > Z_{0.025} = 1.96$ for $\alpha = 0.025$

6) $w_1 = 78$, $\mu_{w_1} = 72$ and $\sigma^2_{w_1} = 108$

$$z_0 = \frac{78 - 72}{10.39} = 0.58$$

Because $Z_0 < 1.96$, fail to reject H_0

7) Conclusion: fail to reject H_0. There is not a significant difference in the heat gain for the heating units at $\alpha = 0.05$.
P-value = 2[1 - P(Z < 0.58)] = 0.5619

10-42 1) The parameters of interest are the mean flight delays

2) $H_0 : \mu_1 = \mu_2$

3) $H_1 : \mu_1 \neq \mu_2$

4) The test statistic is $w_2 = \dfrac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$

5) Reject H_0 if $w \le w^*_{0.01} = 23$. Because $\alpha = 0.01$ and $n_1 = 6$ and $n_2 = 6$, Appendix A, Table X gives the critical value.

6) $w_1 = 40$ and $w_2 = 38$ and because 40 and 38 are greater than 23, fail to reject H_0

7) Conclusion: fail to reject H_0. There is no significant difference in the flight delays at $\alpha = 0.01$.

10-43 a)

1) The parameters of interest are the mean heat gains for heating units

2) $H_0 : \mu_1 = \mu_2$

3) $H_1 : \mu_1 \neq \mu_2$

4) The test statistic is $w_2 = \dfrac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$

6) We reject H_0 if $w \le w^*_{0.01} = 78$. Because $\alpha = 0.01$ and $n_1 = 10$ and $n_2 = 10$, Appendix A, Table X gives the critical value.

7) $w_1 = 77$ and $w_2 = 133$ and because 77 is less than 78, we reject H_0

8) Conclusion: reject H_0 and conclude that there is a significant difference in the heating units at $\alpha = 0.05$.

b)

1) The parameters of interest are the mean heat gain for heating units

2) $H_0 : \mu_1 = \mu_2$

3) $H_1 : \mu_1 \neq \mu_2$

4) The test statistic is $z_0 = \dfrac{W_1 - \mu_{w_1}}{\sigma_{w_1}}$

5) Reject H_0 if $|Z_0| > Z_{0.025} = 1.96$ for $\alpha = 0.05$

6) $w_1 = 77$, $\mu_{w_1} = 105$ and $\sigma^2_{w_1} = 175$

$$z_0 = \frac{77 - 105}{13.23} = -2.12$$

Because $|Z_0| > 1.96$, reject H_0

7) Conclusion: reject H_0 and conclude that there is a difference in the heat gain for the heating units at $\alpha = 0.05$.
P-value =2[1 - P(Z < 2.19)] = 0.034

10-44 a)

1) The parameters of interest are the mean etch rates

2. $H_0 : \mu_1 = \mu_2$

3. $H_1 : \mu_1 \neq \mu_2$

4) The test statistic is $w_2 = \dfrac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$

5) We reject H_0 if $w \le w^*_{0.05} = 78$, because $\alpha = 0.05$ and $n_1 = 10$ and $n_2 = 10$, Appendix A, Table X gives the critical value.

6) w_1 = 73 and w_2 = 137 and because 73 is less than 78, we reject H_0
7) Conclusion: reject H_0 and conclude that there is a significant difference in the mean etch rate at $\alpha = 0.05$.

b)
1) The parameters of interest are the mean temperatures
2) $H_0 : \mu_1 = \mu_2$
3) $H_1 : \mu_1 \neq \mu_2$
4) The test statistic is $z_0 = \dfrac{W_1 - \mu_{w_1}}{\sigma_{w_1}}$
5) We reject H_0 if $|Z_0| > Z_{0.025} = 1.96$ for $\alpha = 0.05$
6) w_1 = 55 , μ_{w_1} =232.5 and $\sigma^2_{w_1}$ =581.25

$$z_0 = \frac{258 - 232.5}{24.11} = 1.06$$

Because $|Z_0| < 1.96$, do not reject H_0
7) Conclusion: fail to reject H_0. There is not a significant difference in the pipe deflection temperatures at $\alpha = 0.05$.
P-value =2[1 - P(Z < 1.06)] = 0.2891

10-45 a)
1) The parameters of interest are the mean temperatures
2) $H_0 : \mu_1 = \mu_2$
3) $H_1 : \mu_1 \neq \mu_2$
4) The test statistic is $w_2 = \dfrac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$
5) Reject H_0 if $w \le w^*_{0.05}$ = 185. Because $\alpha = 0.05$ and $n_1 = 15$ and $n_2 = 15$, Appendix A, Table X gives the critical value.
6) w_1 = 258 and w_2 = 207 and because both 258 and 207 are greater than 185, we fail to reject H_0
7) Conclusion: fail to reject H_0. There is not a significant difference in the mean pipe deflection temperature at α = 0.05.

b)
1) The parameters of interest are the mean etch rates
2) $H_0 : \mu_1 = \mu_2$
3) $H_1 : \mu_1 \neq \mu_2$
4) The test statistic is $z_0 = \dfrac{W_1 - \mu_{w_1}}{\sigma_{w_1}}$
5) We reject H_0 if $|Z_0| > Z_{0.025} = 1.96$ for α=0.05
6) w_1 = 73, μ_{w_1} =105 and $\sigma^2_{w_1}$ =175

$$z_0 = \frac{73 - 105}{13.23} = -2.42$$

Because $|Z_0| > 1.96$, reject H_0
7) Conclusion: reject H_0. There is a significant difference between the mean etch rates.
P-value = P(Z > 2.42 or Z < -2.42) = 0.0155

10-46 a) The data are analyzed in ascending order and ranked as follows:

Group	Distance	Rank
2	244	1
2	258	2
2	260	3
1	263	4.5
2	263	4.5
2	265	6

1	267	7
2	268	8
2	270	9
1	271	11
1	271	11
2	271	11
2	273	13
1	275	14.5
1	275	14.5
1	279	16
2	281	17
1	283	18
1	286	19
1	287	20

The sum of the ranks for group 1 is $w_1 = 135.5$ and for group 2, $w_2 = 74.5$. Because w_2 is less than $w_{0.05} = 78$, we reject the null hypothesis that both groups have the same mean.

b) When the sample sizes are equal it does not matter which group we select for w_1

$$\mu_{W_1} = \frac{10(10+10+1)}{2} = 105$$

$$\sigma^2_{W_1} = \frac{10*10(10+10+1)}{12} = 175$$

$$Z_0 = \frac{135.5-105}{\sqrt{175}} = 2.31$$

Because $z_0 > z_{0.025} = 1.96$, reject H_0 and conclude that the sample means for the two groups are different.
When $z_0 = 2.31$, P-value = P(Z < -2.31 or Z > 2.31) = 0.021

10-47

upper group	lower group
17.8	11.8
23.3	16.4
23.4	18.4
23.8	26
31	33.3
33.6	34.1
41.5	35.8
43.9	38.7
48.9	41.8
49.5	43.1
56	47.3
56.4	55
65	59.6
75.8	
78.8	

Count	3	2
Exceedences (E)	3+2 = 5	

Because the exceedences E = 5 < 7, do not reject the hypothesis at 0.05 level of significance and conclude that the rivers do not differ significantly in the mean algae content.

10-48

If heats 5 and 6 are combined both the minimum and maximum value occurs in the group with heats 5 and 6. Therefore, Tukey's test does not apply.

Consider heat 5 versus heat 7

Upper Group	Heat	Lower Group	Heat
49.02	5	48.54	7
49.49	5	48.67	7
49.6	5	48.93	7
49.78	5	48.93	7
49.95	5	48.97	7
50.08	5	49.03	7
		49.29	7

The number of exceedances is 5 + 5 = 10 ≥ 7. Therefore there is a significant difference in times at $\alpha = 0.05$. This agrees with the result from the t test in the previous exercise.
Because the critical value for the number of exceedances is 10 at $\alpha = 0.05$, there is a significant difference in times even at $\alpha = 0.01$.

Consider heat 6 versus heat 7

Lower Group	Heat	Upper Group	Heat
48.19	6	48.54	7
48.29	6	48.67	7
48.54	6	48.93	7
48.6	6	48.93	7
48.67	6	48.97	7
49.18	6	49.03	7
49.2	6	49.29	7

The number of exceedances is 2.5 + 1 = 3.5 < 7. Therefore there is no significant difference in times at $\alpha = 0.05$. This agrees with the result from the t test in the previous exercise.

<u>Section 10-4</u>

10-49 a) $\bar{d} = 0.2738$ $s_d = 0.1351$, n = 9

95% confidence interval:

$$\bar{d} - t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right) \le \mu_d \le \bar{d} + t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right)$$

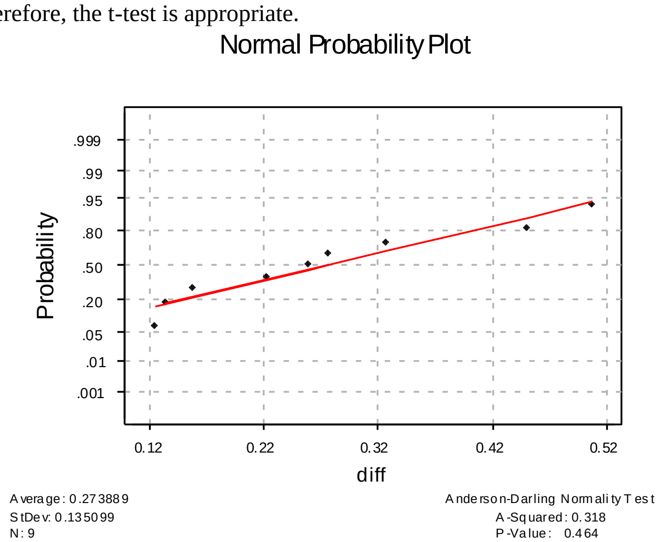
$$0.2738 - 2.306\left(\frac{0.1351}{\sqrt{9}}\right) \le \mu_d \le 0.2738 + 2.306\left(\frac{0.1351}{\sqrt{9}}\right)$$

$0.1699 \le \mu_d \le 0.3776$

With 95% confidence, the mean shear strength of Karlsruhe method exceeds the mean shear strength of the Lehigh method by between 0.1699 and 0.3776. Because zero is not included in this interval, the interval is consistent with rejecting the null hypothesis that the means are equal.

The 95% confidence interval is directly related to a test of hypothesis with 0.05 level of significance and the conclusions reached are identical.

b) It is only necessary for the differences to be normally distributed for the paired *t*-test to be appropriate and reliable. Therefore, the t-test is appropriate.



10-50 a)

1) The parameter of interest is the difference between the mean parking times, μ_d.

2) $H_0: \mu_d = 0$

3) $H_1: \mu_d \ne 0$

4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

5) Reject the null hypothesis if $t_0 < -t_{0.05,13}$ where $-t_{0.05,13} = -1.771$ or $t_0 > t_{0.05,13}$ where $t_{0.05,13} = 1.771$ for $\alpha = 0.10$

6) $\bar{d} = 1.21$

$s_d = 12.68$

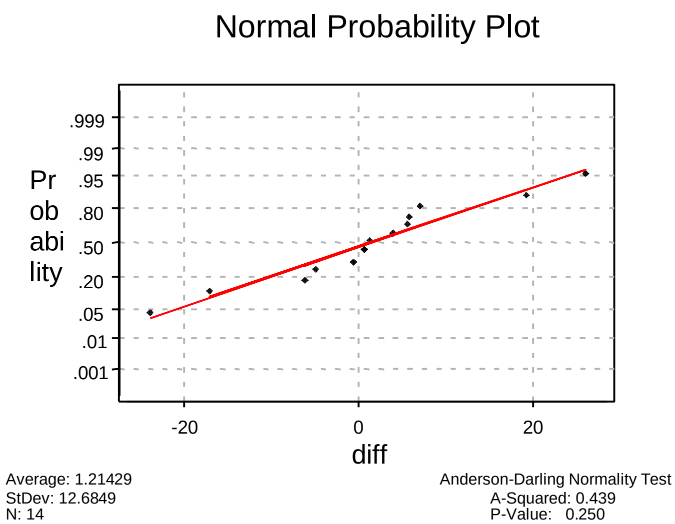
$n = 14$

$$t_0 = \frac{1.21}{12.68/\sqrt{14}} = 0.357$$

7) Conclusion: Because $-1.771 < 0.357 < 1.771$, fail to reject the null. The data fail to support the claim that the two cars have different mean parking times at the 0.10 level of significance.

b) The result is consistent with the confidence interval constructed because zero is included in the 90% confidence interval.

c) The data fall approximately along a line in the normal probability plots. Therefore, the assumption of normality does not appear to be violated.



10-51 $\bar{d}$ = 868.375 s_d = 1290, n = 8 where d_i = brand 1 - brand 2

99% confidence interval:

$$\bar{d} - t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right) \le \mu_d \le \bar{d} + t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right)$$

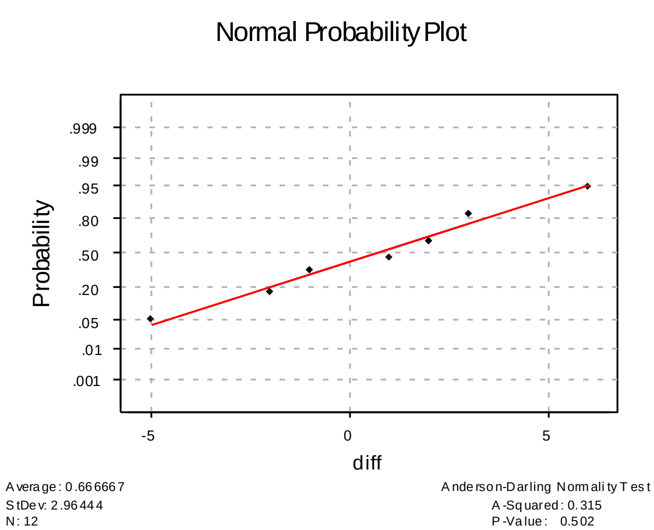
$$868.375 - 3.499\left(\frac{1290}{\sqrt{8}}\right) \le \mu_d \le 868.375 + 3.499\left(\frac{1290}{\sqrt{8}}\right)$$

$-727.46 \le \mu_d \le 2464.21$

Because this confidence interval contains zero, there is no significant difference between the two brands of tire at a 1% significance level.

10-52

a) The data fall approximately along a line in the normal probability plots. Therefore, the assumption of normality does not appear to be violated.



b) $\bar{d}$ = 0.667 s_d = 2.964, n = 12

95% confidence interval:

$$\bar{d} - t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right) \le \mu_d \le \bar{d} + t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right)$$
$$0.667 - 2.201\left(\frac{2.964}{\sqrt{12}}\right) \le \mu_d \le 0.667 + 2.201\left(\frac{2.964}{\sqrt{12}}\right)$$
$$-1.216 \le \mu_d \le 2.55$$

Because zero is contained within this interval, one cannot conclude that one design language is preferable at a 5% significance level

10-53 a)

1) The parameter of interest is the difference in blood cholesterol level, μ_d where d_i = Before − After.
2) $H_0: \mu_d = 0$
3) $H_1: \mu_d > 0$
4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

5) Reject the null hypothesis if $t_0 > t_{0.05,14}$ where $t_{0.05,14} = 1.761$ for $\alpha = 0.05$
6) $\bar{d} = 26.867$
$s_d = 19.04$
$n = 15$

$$t_0 = \frac{26.867}{19.04/\sqrt{15}} = 5.465$$

7) Conclusion: Because 5.465 > 1.761, reject the null hypothesis. The data support the claim that the mean difference in cholesterol levels is significantly less after diet and an aerobic exercise program at the 0.05 level of significance.

P-value = $P(t > 5.565) \cong 0$

b) 95% confidence interval:

$$\bar{d} - t_{\alpha,n-1}\left(\frac{s_d}{\sqrt{n}}\right) \le \mu_d$$
$$26.867 - 1.761\left(\frac{19.04}{\sqrt{15}}\right) \le \mu_d$$
$$18.20 \le \mu_d$$

Because the lower bound is positive, the mean difference in blood cholesterol level is significantly less after the diet and aerobic exercise program.

10-54 a)

1) The parameter of interest is the mean difference in natural vibration frequencies, μ_d where d_i = finite element – equivalent plate.
2) $H_0: \mu_d = 0$
3) $H_1: \mu_d \ne 0$
4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

5) Reject the null hypothesis if $t_0 < -t_{0.025,6}$ or $t_0 > t_{0.025,6}$ where $t_{0.005,6} = 2.447$ for $\alpha = 0.05$
6) $\bar{d} = -5.49$
$s_d = 5.924$
$n = 7$

$$t_0 = \frac{-5.49}{5.924/\sqrt{7}} = -2.45$$

7) Conclusion: Because −2.45 < −2.447, reject the null hypothesis. The two methods have different mean values for natural vibration frequency at the 0.05 level of significance.

b) 95% confidence interval:

$$\bar{d} - t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right) \le \mu_d \le \bar{d} + t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right)$$

$$-5.49 - 2.447\left(\frac{5.924}{\sqrt{7}}\right) \le \mu_d \le -5.49 + 2.447\left(\frac{5.924}{\sqrt{7}}\right)$$

$$-10.969 \le \mu_d \le -0.011$$

With 95% confidence, the mean difference between the natural vibration frequency from the equivalent plate method and the finite element method is between −10.969 and −0.011 cycles.

10-55 a)

1) The parameter of interest is the difference in mean weight, μ_d , where d_i = weight before − weight after.

2) $H_0: \mu_d = 0$

3) $H_1: \mu_d > 0$

4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

5) Reject the null hypothesis if $t_0 > t_{0.05,9}$ where $t_{0.05,9} = 1.833$ for $\alpha = 0.05$

6) $\bar{d} = 17$

$s_d = 6.41$

$n = 10$

$$t_0 = \frac{17}{6.41/\sqrt{10}} = 8.387$$

7) Conclusion: Because 8.387 > 1.833, reject the null hypothesis and conclude that the mean weight loss is significantly greater than zero. That is, the data support the claim that this particular diet modification program is effective in reducing weight at the 0.05 level of significance.

b)

1) The parameter of interest is the difference in mean weight loss, μ_d, where d_i = Before − After.

2) $H_0: \mu_d = 10$

3) $H_1: \mu_d > 10$

4) The test statistic is

$$t_0 = \frac{\bar{d} - \Delta_0}{s_d/\sqrt{n}}$$

5) Reject the null hypothesis if $t_0 > t_{0.05,9}$ where $t_{0.05,9} = 1.833$ for $\alpha = 0.05$

6) $\bar{d} = 17$

$s_d = 6.41$

$n = 10$

$$t_0 = \frac{17-10}{6.41/\sqrt{10}} = 3.45$$

7) Conclusion: Because 3.45 > 1.833, reject the null hypothesis. There is evidence to support the claim that this particular diet modification program is effective in producing a mean weight loss of at least 10 lbs at the 0.05 level of significance.

c) Use s_d as an estimate for σ:

$$n = \left(\frac{(z_\alpha + z_\beta)\sigma_d}{10}\right)^2 = \left(\frac{(1.645+1.29)6.41}{10}\right)^2 = 3.53, \quad n = 4$$

Yes, the sample size of 10 is adequate for this test.

10-56 a)

1) The parameter of interest is the mean difference in impurity level, μ_d, where d_i = Test 1 – Test 2.

2) $H_0: \mu_d = 0$

3) $H_1: \mu_d \neq 0$

4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

5) Reject the null hypothesis if $t_0 < -t_{0.005,7}$ or $t_0 > t_{0.005,7}$ where $t_{0.005,7}$ = 3.499 for α = 0.01

6) $\bar{d} = -0.2125$
$s_d = 0.1727$
$n = 8$

$$t_0 = \frac{-0.2125}{0.1727/\sqrt{8}} = -3.48$$

7) Conclusion: Because –3.499 < –3.48 < 3.499, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the tests generate different mean impurity levels at α = 0.01.

b)

1) The parameter of interest is the mean difference in impurity level, μ_d, where d_i = Test 1 – Test 2.

2) $H_0: \mu_d + 0.1 = 0$

3) $H_1: \mu_d + 0.1 < 0$

4) The test statistic is

$$t_0 = \frac{\bar{d} + 0.1}{s_d / \sqrt{n}}$$

5) Reject the null hypothesis if $t_0 < -t_{0.05,7}$ where $t_{0.05,7}$ = 1.895 for α = 0.05

6) $\bar{d} = -0.2125$
$s_d = 0.1727$
$n = 8$

$$t_0 = \frac{-0.2125 + 0.1}{0.1727/\sqrt{8}} = -1.8424$$

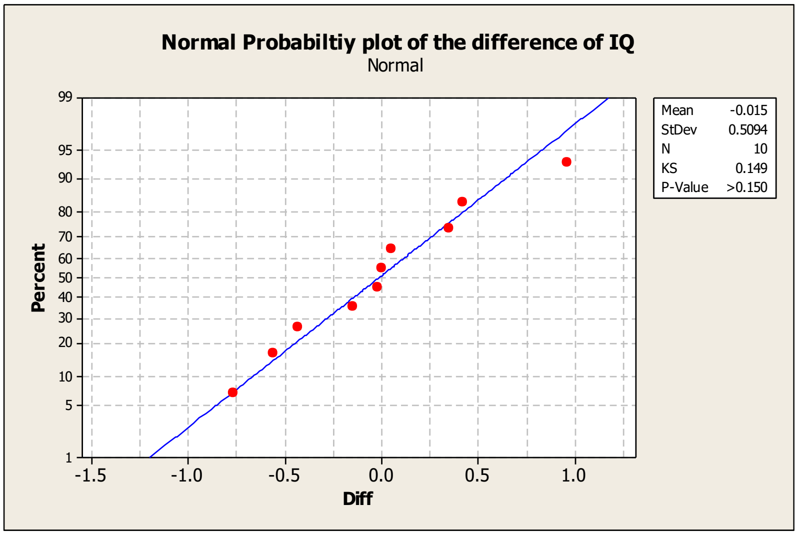
7) Conclusion: Because –1.895 < –1.8424, fail to reject the null hypothesis at the 0.05 level of significance.

c) β = 1 - 0.9 = 0.1

$$d = \frac{|0.1|}{0.1727} = 0.579$$

n = 8 is not an adequate sample size. From the chart VII*g*, n ≈ 30

10-57 a) The data in the probability plot fall approximately along a line. Therefore, the normality assumption is reasonable.



b)
$\bar{d} = -0.015$
$s_d = 0.5093$
$n = 10$
$t_{0.025,9} = 2.262$

95% confidence interval:
$$\bar{d} - t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right) \le \mu_d \le \bar{d} + t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right)$$
$-0.379 \le \mu_d \le 0.3493$

Because zero is contained in the confidence interval, there is not sufficient evidence that the mean IQ depends on birth order.

c) β = 1- 0.9 = 0.1
$$d = d = \frac{|\delta|}{\sigma} = \frac{1}{s_d} = 1.96$$
Thus $6 \le n$ would be enough.

10-58 a) Let $x_{12} = x_2 - x_1$ and $x_{23} = x_3 - x_2$ and $x_d = x_{23} - x_{12}$

1) The parameter of interest is the mean difference in circumference μ_d where $x_d = x_{23} - x_{12}$

2) $H_0: \mu_d = 0$

3) $H_1: \mu_d \ne 0$

4) The test statistic is
$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

5) Reject the null hypothesis if $t_0 < -t_{0.05,4}$ or $t_0 > t_{0.05,4}$ where $t_{0.05,4} = 2.132$ for $\alpha = 0.10$

6) $\bar{d} = 8.6$

$s_d = 7.829$

$n = 5$ $$t_0 = \frac{8.6}{7.829/\sqrt{5}} = 2.456$$

7) Conclusion: Because 2.132 < 2.456, reject the null hypothesis. The means are significantly different at α = 0.1.

b) Let $x_{67} = x_7 - x_6$

Let $x_d = x_{12} - x_{67}$

1) The parameter of interest is the mean difference in circumference μ_d where $x_d = x_{12} - x_{67}$

2) H_0 : $\mu_d = 0$

3) H_1 : $\mu_d > 0$

4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

5) Reject the null hypothesis if $t_0 > t_{0.05,4}$ where $t_{0.05,4}$ = 2.132 for α = 0.05

6) $\bar{d}$ = -24.4
$s_d = 7.5$

$n = 5$ $$t_0 = \frac{-24.4}{7.5/\sqrt{5}} = 7.27$$

7) Conclusion: Because 7.27 > 2.132, reject the null hypothesis. The means are significantly different at α = 0.1.

P-value = P(t > 7.27) ≈ 0

c) No, the paired t test uses the differences to conduct the inference.

10-59 1) Parameters of interest are the median cholesterol levels for two activities.

2) $H_0 : \tilde{\mu}_D = 0$ or 2) $H_0 : \tilde{\mu}_1 - \tilde{\mu}_2 = 0$

3) $H_1 : \tilde{\mu}_D > 0$ 3) $H_1 : \tilde{\mu}_1 - \tilde{\mu}_2 > 0$

4) r^-

5) Because α = 0.05 and n = 15, Appendix A, Table VIII gives the critical value of $r^*_{0.05} = 3$. We reject H_0 in favor of H_1 if $r^- \le 3$.

6) The test statistic is $r^- = 2$.

Observation	Before	After	Difference	Sign
1	265	229	36	+
2	240	231	9	+
3	258	227	31	+
4	295	240	55	+
5	251	238	13	+
6	245	241	4	+
7	287	234	53	+
8	314	256	58	+
9	260	247	13	+
10	279	239	40	+
11	283	246	37	+
12	240	218	22	+
13	238	219	19	+
14	225	226	-1	-

15	247	233	14	+

P-value = P(R$^+$ $\geq r^+$ = 14 | p = 0.5) = $\sum_{r=13}^{15} \binom{15}{r} (0.5)^r (0.5)^{20-r} = 0.00049$

7) Conclusion: Because the P-value = 0.00049 is less than α = 0.05, reject the null hypothesis. There is a significant difference in the median cholesterol levels after diet and exercise at α = 0.05.

10-60 1) The parameters of interest are the median cholesterol levels for two activities.

2) and 3) $H_0 : \mu_D = 0$ or $H_0 : \mu_1 - \mu_2 = 0$
$H_1 : \mu_D > 0$ $H_1 : \mu_1 - \mu_2 > 0$

4) w^-

5) Reject H_0 if $w^- \leq w^*_{0.05, n=15} = 30$ for α = 0.05

6) The sum of the negative ranks is w^- = (1+15) = 16.

Observation	Before	After	Difference	Signed Rank
14	225	226	-1	-1
2	240	231	9	3
5	251	238	13	4.5
9	260	247	13	4.5
15	247	233	14	6
13	238	219	19	7
12	240	218	22	8
3	258	227	31	9
1	265	229	36	10
11	283	246	37	11
10	279	239	40	12
7	287	234	53	13
4	295	240	55	14
8	314	256	58	15
6	245	241	4	2

7) Conclusion: Because w^- = 1 is less than the critical value $w^*_{0.05, n=15} = 30$, reject the null hypothesis. There is a significant difference in the mean cholesterol levels after diet and exercise at α = 0.05.

The previous exercise tests the difference in the median cholesterol levels after diet and exercise while this exercise tests the difference in the mean cholesterol levels after diet and exercise.

10-61

a) No. There are no pairs in this data.
b) The appropriate test is a two-sample t test.
1) The means of the nonconfined and confined groups are μ_1 and μ_2, respectively
2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) H_1 : $\mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$. The alternative is two-sided.
4) The test statistic is (assume equal variances)

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ where $-t_{0.025,18} = -2.101$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$ where $t_{0.025,18} = 2.101$ for α = 0.05

6) $\bar{x}_1 = 10.58$ $\bar{x}_2 = 9.78$ $s_p = \sqrt{\dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

$s_1 = 0.459$ $s_2 = 0.598$ $= \sqrt{\dfrac{9(0.459)^2+9(0.598)^2}{18}} = 0.533$

$n_1 = 10$ $n_2 = 10$

$$t_0 = \frac{(10.58-9.78)}{0.533\sqrt{\frac{1}{10}+\frac{1}{10}}} = 3.36$$

7) Conclusion: Because 2.101 < 3.36, reject the null hypothesis. The brain wave activity means differ significantly at $\alpha = 0.05$.

10-62

a) This is a one sided test.

b)

1) The parameter of interest is the mean difference in fluency, μ_d

2) $H_0: \mu_d = 0$

3) $H_1: \mu_d > 0$

4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

5) Reject the null hypothesis if $t_0 > t_{0.05,98}$ where $t_{0.05,98} = 1.661$ for $\alpha = 0.05$.

6) $\bar{d} = 1.07$

$s_d = 3.195$

$n = 99$

$$t_0 = \frac{1.07}{3.195/\sqrt{99}} = 3.332$$

7) Conclusion: Because 3.332 > 1.661, reject the null hypothesis. The mean difference is positive.

c) Because the tests are conducted for every subject independently before and after joining the study, and the results are available in pairs for the people, this can be viewed as a paired t-test.

d) The significant result implies that even with no medication (placebo) the recall improved, possibly because subjects became more familiar with the test. Consequently, it is important to include an untreated (placebo) group to evaluate the effect of gingko.

Section 10-5

10-63 a) $f_{0.25,5,10} = 1.59$

b) $f_{0.10,24,9} = 2.28$

c) $f_{0.05,8,15} = 2.64$

d) $f_{0.75,5,10} = \dfrac{1}{f_{0.25,10,5}} = \dfrac{1}{1.89} = 0.529$

e) $f_{0.90,24,9} = \dfrac{1}{f_{0.10,9,24}} = \dfrac{1}{1.91} = 0.525$

f) $f_{0.95,8,15} = \dfrac{1}{f_{0.05,15,8}} = \dfrac{1}{3.22} = 0.311$

10-64 a) $f_{0.25,7,15} = 1.47$

d) $f_{0.75,7,15} = \dfrac{1}{f_{0.25,15,7}} = \dfrac{1}{1.68} = 0.596$

b) $f_{0.10,10,12} = 2.19$

c) $f_{0.01,20,10} = 4.41$

e) $f_{0.90,10,12} = \dfrac{1}{f_{0.10,12,10}} = \dfrac{1}{2.28} = 0.438$

f) $f_{0.99,20,10} = \dfrac{1}{f_{0.01,10,20}} = \dfrac{1}{3.37} = 0.297$

10-65 1) The parameters of interest are the standard deviations σ_1, σ_2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 < \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{.0.95,4,9} = 1/f_{.0.05,9,4} = 1/6 = 0.1666$ for $\alpha = 0.05$

6) $n_1 = 5$ $n_2 = 10$ $s_1^2 = 23.2$ $s_2^2 = 28.8$

$$f_0 = \frac{(23.2)}{(28.8)} = 0.806$$

7) Conclusion: Because 0.1666 < 0.806 do not reject the null hypothesis.

95% confidence interval:

$$\frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha, n_2-1, n_1-1}$$

$\dfrac{\sigma_1^2}{\sigma_2^2} \le (\dfrac{23.2}{28.8}) f_{0.05,9,4}$ where $f_{.05,9,4} = 6.00$ $\dfrac{\sigma_1^2}{\sigma_2^2} \le 4.83$ or $\dfrac{\sigma_1}{\sigma_2} \le 2.20$

Because the value one is contained within this interval, there is no significant difference in the variances.

10-66 1) The parameters of interest are the standard deviations, σ_1, σ_2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 > \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 > f_{.0.01,19,7} \cong 6.16$ for $\alpha = 0.01$

6) $n_1 = 20$ $n_2 = 8$ $s^2{}_1 = 4.5$ $s_2{}^2 = 2.3$

$$f_0 = \frac{4.5}{2.3} = 1.956$$

7) Conclusion: Because 6.16 > 1.956, fail to reject the null hypothesis.

95% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{0.99, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2}$$

$1.956(1/6.16) \le \dfrac{\sigma_1^2}{\sigma_2^2}$ $0.318 \le \dfrac{\sigma_1^2}{\sigma_2^2}$

Because the value one is contained within this interval, there is no significant difference in the variances.

10-67 a)

1) The parameters of interest are the standard deviations, σ_1, σ_2

2) H0 : $\sigma_1^2 = \sigma_2^2$

3) H1 : $\sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if f0 < $f_{.0.975,14,14}$ = 0.33 or f0 > $f_{0.025,14,14}$ = 3 for α = 0.05

6) $n_1 = 15$ $n_2 = 15$ $s_1{}^2 = 2.3$ $s_2{}^2 = 1.9$

$$f_0 = \frac{2.3}{1.9} = 1.21$$

7) Conclusion: Because 0.333 < 1.21, < 3 fail to reject the null hypothesis. There is not sufficient evidence that there is a difference in the standard deviations.

95% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha/2,n_2-1,n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \left(\frac{s_1^2}{s_2^2}\right) f_{\alpha/2,n_2-1,n_1-1}$$

$$(1.21)0.333 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq (1.21)3 \qquad 0.403 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 3.63$$

Because the value one is contained within this interval, there is no significant difference in the variances.

b) $\lambda = \dfrac{\sigma_1}{\sigma_2} = 2$

$n_1 = n_2 = 15$

α = 0.05

Chart VII (o) we find β = 0.35 then the power 1- β = 0.65

c) β = 0.05 and $\sigma_2 = \sigma_1 / 2$ so that $\dfrac{\sigma_1}{\sigma_2} = 2$ and n ≈ 31

10-68 1) The parameters of interest are the variances of concentration, σ_1^2, σ_2^2

2) H0 : $\sigma_1^2 = \sigma_2^2$

3) H1 : $\sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if f0 < $f_{0.975,9,15}$ where $f_{0.975,9,15}$ = 0.265 or f0 > $f_{0.025,9,15}$ where $f_{0.025,9,15}$ = 3.12 for α = 0.05

6) $n_1 = 10$ $n_2 = 16$

$s_1 = 4.7$ $s_2 = 5.8$

$$f_0 = \frac{(4.7)^2}{(5.8)^2} = 0.657$$

7) Conclusion: Because 0.265 < 0.657 < 3.12, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the two population variances differ at the 0.05 level of significance.

10-69

a)

1) The parameters of interest are the time to assemble standard deviations, σ_1, σ_2 where Group 1 = men and Group 2 = women

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{1-\alpha/2, n_1-1, n_2-1} = 0.365$ or $f_0 > f_{\alpha/2, n_1-1, n_2-1} = 2.86$ for $\alpha = 0.02$

6) $n_1 = 25$ $n_2 = 21$ $s_1 = 0.98$ $s_2 = 1.02$

$$f_0 = \frac{(0.98)^2}{(1.02)^2} = 0.923$$

7) Conclusion: Because 0.365 < 0.923 < 2.86, fail to reject the null hypothesis. There is not sufficient evidence to support the claim that men and women differ in repeatability for this assembly task at the 0.02 level of significance.

ASSUMPTIONS: Assume random samples from two normal distributions.

b) 98% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha/2, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right) f_{\alpha/2, n_2-1, n_1-1}$$

$$f_{1-\alpha/2, n_2-1, n_1-1} = \frac{1}{f_{\alpha/2, n_1-1, n_2-1}} = \frac{1}{f_{0.01,24,20}} = \frac{1}{2.86} = 0.350$$

$$(0.923)0.350 \le \frac{\sigma_1^2}{\sigma_2^2} \le (0.923)2.73$$

$$0.323 \le \frac{\sigma_1^2}{\sigma_2^2} \le 2.527$$

Because the value one is contained within this interval, there is no significant difference between the variance of the repeatability of men and women for the assembly task at a 2% significance level.

10-70 a) 90% confidence interval for the ratio of variances:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha/2, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right) f_{\alpha/2, n_2-1, n_1-1}$$

$$\left(\frac{0.6^2}{0.8^2}\right)0.156 \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{0.6^2}{0.8^2}\right)6.39 \qquad 0.08775 \le \frac{\sigma_1^2}{\sigma_2^2} \le 3.594$$

b) 95% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha/2, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right) f_{\alpha/2, n_2-1, n_1-1}$$

$$\left(\frac{(0.6)^2}{(0.8)^2}\right)0.104 \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{(0.6)^2}{(0.8)^2}\right)9.60 \qquad 0.0585 \le \frac{\sigma_1^2}{\sigma_2^2} \le 5.4$$

The 95% confidence interval is wider than the 90% confidence interval.

c) 90% lower-sided confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2}$$

$$\left(\frac{(0.6)^2}{(0.8)^2}\right)0.243 \le \frac{\sigma_1^2}{\sigma_2^2} \qquad 0.137 \le \frac{\sigma_1^2}{\sigma_2^2}$$

A 90% lower confidence bound on $\frac{\sigma_1}{\sigma_2}$ is given by $0.370 \le \frac{\sigma_1}{\sigma_2}$

10-71 a) 90% confidence interval for the ratio of variances:

$$\left(\frac{s_1^2}{s_2^2}\right)f_{1-\alpha/2,n_2-1,n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right)f_{\alpha/2,n_2-1,n_1-1}$$

$$\left(\frac{0.35}{0.40}\right)0.421 \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{0.35}{0.40}\right)2.445 \quad 0.369 \le \frac{\sigma_1^2}{\sigma_2^2} \le 2.139 \quad 0.607 \le \frac{\sigma_1}{\sigma_2} \le 1.463$$

b) 95% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right)f_{1-\alpha/2,n_2-1,n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right)f_{\alpha/2,n_2-1,n_1-1}$$

$$\left(\frac{0.35}{0.40}\right)0.342 \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{0.35}{0.40}\right)2.82 \quad 0.311 \le \frac{\sigma_1^2}{\sigma_2^2} \le 2.558 \qquad 0.557 \le \frac{\sigma_1}{\sigma_2} \le 1.599$$

The 95% confidence interval is wider than the 90% confidence interval.

c) 90% lower-sided confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right)f_{1-\alpha,n_2-1,n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2}$$

$$\left(\frac{0.35}{0.40}\right)0.512 \le \frac{\sigma_1^2}{\sigma_2^2} \qquad 0.448 \le \frac{\sigma_1^2}{\sigma_2^2} \qquad 0.669 \le \frac{\sigma_1}{\sigma_2}$$

10-72 1) The parameters of interest are the strength variances, σ_1^2, σ_2^2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \ne \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.975,9,15}$ where $f_{0.975,9,15} = 0.265$ or $f_0 > f_{0.025,9,15}$ where $f_{0.025,9,15} = 3.12$ for $\alpha = 0.05$

6) $n_1 = 10$ $\qquad n_2 = 16$

$s_1 = 12 \quad s_2 = 22$

$$f_0 = \frac{(12)^2}{(22)^2} = 0.297$$

7) Conclusion: Because 0.265 < 0.297 < 3.12, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the population variances differ at the 0.05 level of significance.

10-73 1) The parameters of interest are the melting variances, σ_1^2, σ_2^2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \ne \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.975,20,20}$ where $f_{0.975,20,20} = 0.4058$ or $f_0 > f_{0.025,20,20}$ where $f_{0.025,20,20} = 2.46$ for $\alpha = 0.05$

6) $n_1 = 21$ $n_2 = 21$
$s_1 = 4$ $s_2 = 3$

$$f_0 = \frac{(4)^2}{(3)^2} = 1.78$$

7) Conclusion: Because 0.4058 < 1.78 < 2.46, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the population variances differ at the 0.05 level of significance.

10-74 1) The parameters of interest are the thickness variances, σ_1^2, σ_2^2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.995,10,12}$ where $f_{0.995,10,12} = 0.1766$ or $f_0 > f_{0.005,10,12}$ where $f_{0.005,10,12} = 5.0855$ for $\alpha = 0.01$

6) $n_1 = 11$ $n_2 = 13$
$s_1 = 10.2$ $s_2 = 20.1$

$$f_0 = \frac{(10.2)^2}{(20.1)^2} = 0.2575$$

7) Conclusion: Because 0.1766 < 0.2575 < 5.0855, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the thickness variances differ at the 0.01 level of significance.

10-75 1) The parameters of interest are the overall distance standard deviations, σ_1, σ_2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{.0.975,9,9} = 0.248$ or $f_0 > f_{0.025,9,9} = 4.03$ for $\alpha = 0.05$

6) $n_1 = 10$ $n_2 = 10$ $s_1 = 8.03$ $s_2 = 10.04$

$$f_0 = \frac{(8.03)^2}{(10.04)^2} = 0.640$$

7) Conclusion: Because 0.248 < 0.640 < 4.04, fail to reject the null hypothesis. There is not sufficient evidence that the standard deviations of the overall distances of the two brands differ at the 0.05 level of significance.

95% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha/2, n_2-1, n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \left(\frac{s_1^2}{s_2^2}\right) f_{\alpha/2, n_2-1, n_1-1}$$

$$(0.640)0.248 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq (0.640)4.03 \qquad 0.159 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 2.579$$

A 95% lower confidence bound on the ratio of standard deviations is given by $0.399 \leq \frac{\sigma_1}{\sigma_2} \leq 1.606$

Because the value one is contained within this interval, there is no significant difference in the variances of the distances at a 5% significance level.

10-76 1) The parameters of interest are the time to assemble standard deviations, σ_1, σ_2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{.0.975,11,11} = 0.288$ or $f_0 > f_{0.025,11,11} = 3.474$ for $\alpha = 0.05$

6) $n_1 = 12$ $n_2 = 12$ $s_1 = 0.0217$ $s_2 = 0.0175$

$$f_0 = \frac{(0.0217)^2}{(0.0175)^2} = 1.538$$

7) Conclusion: Because 0.288 < 1.538 < 3.474, fail to reject the null hypothesis. There is not sufficient evidence that there is a difference in the standard deviations of the coefficients of restitution between the two clubs at the 0.05 level of significance.

95% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha/2, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right) f_{\alpha/2, n_2-1, n_1-1}$$

$$(1.538)0.288 \le \frac{\sigma_1^2}{\sigma_2^2} \le (1.538)3.474 \qquad 0.443 \le \frac{\sigma_1^2}{\sigma_2^2} \le 5.343$$

A 95% lower confidence bound the ratio of standard deviations is given by $0.666 \le \frac{\sigma_1}{\sigma_2} \le 2.311$

Because the value one is contained within this interval, there is no significant difference in the variances of the coefficient of restitution at a 5% significance level.

10-77 1) The parameters of interest are the variances of the weight measurements between the two sheets of paper, σ_1^2, σ_2^2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{.0.975,14,14} = 0.33$ or $f_0 > f_{0.025,14,14} = 3$ for $\alpha = 0.05$

6) $n_1 = 15$ $n_2 = 15$ $s_1^{\ 2} = 0.00831^2$ $s_2^{\ 2} = 0.00714^2$

$$f_0 = 1.35$$

7) Conclusion: Because 0.333 < 1.35 < 3, fail to reject the null hypothesis. There is not sufficient evidence that there is a difference in the variances of the weight measurements between the two sheets of paper at $\alpha = 0.05$.

95% confidence interval:

$$\left(\frac{s_1^2}{s_2^2}\right) f_{1-\alpha/2, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{s_1^2}{s_2^2}\right) f_{\alpha/2, n_2-1, n_1-1}$$

$$(1.35)0.333 \le \frac{\sigma_1^2}{\sigma_2^2} \le (1.35)3 \qquad 0.45 \le \frac{\sigma_1^2}{\sigma_2^2} \le 4.05$$

Because the value one is contained within this interval, there is no significant difference in the variances.

10-78 a)

1) The parameters of interest are the thickness variances, σ_1^2, σ_2^2

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.99,7,7}$ where $f_{0.99,7,7} = 0.143$ or $f_0 > f_{0.01,7,7}$ where $f_{0.01,7,7} = 6.99$ for $\alpha = 0.02$

6) $n_1 = 8$ $n_2 = 8$
$s_1 = 0.11$ $s_2 = 0.09$

$$f_0 = \frac{(0.11)^2}{(0.09)^2} = 1.49$$

7) Conclusion: Because 0.143 < 1.49 < 6.99, fail to reject the null hypothesis. The thickness variances do not significantly differ at the 0.02 level of significance.

b) If one population standard deviation is to be 50% larger than the other, then $\lambda = 2$. Using $n = 8$, $\alpha = 0.01$ and Chart VII (p), we obtain $\beta \cong 0.85$. Therefore, $n = n_1 = n_2 = 8$ is not adequate to detect this difference with high probability.

10-79 a)

1) The parameters of interest are the etch-rate variances, σ_1^2, σ_2^2 .

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.975,9,9} = 0.248$ or $f_0 > f_{0.025,9,9} = 4.03$ for $\alpha = 0.05$

6) $n_1 = 10$ $n_2 = 10$
$s_1 = 0.422$ $s_2 = 0.231$

$$f_0 = \frac{(0.422)^2}{(0.231)^2} = 3.337$$

7) Conclusion: Because 0.248 < 3.337 < 4.03, fail to reject the null hypothesis. There is not sufficient evidence that the etch rate variances differ at the 0.05 level of significance.

b) With $\lambda = \sqrt{2} = 1.4\, \beta = 0.10$ and $\alpha = 0.05$, we find from Chart VII*o* that $n_1^* = n_2^* = 100$. Therefore, samples of size 10 would not be adequate.

10-80

1) The parameters of interest are the swim time variances, σ_1^2, σ_2^2 .

2) $H_0: \sigma_1^2 = \sigma_2^2$

3) $H_1: \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if the P-value is less than $\alpha = 0.05$

F-Test Two-Sample for Variances for Heats 5 and 7

	Variable 1	*Variable 2*
Mean	49.65333	48.90857
Variance	0.143347	0.059614
Observations	6	7
df	5	6
F	2.404569	
P(F<=f)	0.1578	
F Critical	4.387374	

Conclusion: The *P*-value = 0.16. There is not sufficient evidence that the variances differ at the 0.05 level of significance.

F-Test Two-Sample for Variances for Heats 6 and 7

	Variable 1	*Variable 2*
Mean	48.66714	48.90857
Variance	0.156257	0.059614
Observations	7	7
df	6	6
F	2.621136	
P(F<=f	0.133001	
F Critical	4.283866	

Conclusion: *P*-value = 0.13. Because the P-value > 0.05, there is not sufficient evidence that the variances differ at the 0.05 level of significance.

10-81

1) The parameters of interest are the algae concentration variances, σ_1^2, σ_2^2 .

2) $H_0 : \sigma_1^2 = \sigma_2^2$

3) $H_1 : \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.975,14,12} = 1/3.05=0.328$ or $f_0 > f_{0.025,14,12} = 3.05$ for $\alpha = 0.05$

6) $n_1 = 15$ $\quad n_2 = 13$
$s_1 = 19.348$ $\quad s_2 = 14.503$

$$f_0 = \frac{(19.348)^2}{(14.503)^2} = 1.78$$

7) Conclusion: Because 0.328 < 1.78 < 3.05, fail to reject the null hypothesis. There is not sufficient evidence that the algae concentration variances differ at the 0.05 level of significance.

Section 10-6

10-82 a) This is a two-sided test because the hypotheses are $p_1 - p_2 = 0$ versus not equal to 0.

b) $\hat{p}_1 = \frac{54}{250} = 0.216$ $\hat{p}_2 = \frac{60}{290} = 0.207$ $\hat{p} = \frac{54+60}{250+290} = 0.2111$

Test statistic is $z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$$z_0 = \frac{0.0091}{\sqrt{(0.2111)(1-0.2111)\left(\frac{1}{250}+\frac{1}{290}\right)}} = 0.2584$$

P-value = 2[1 – P(Z < 0.2584)] = 2[1-0.6020] = 0.796

c) Because the P-value is greater than α = 0.05, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the proportions differ at the 0.05 level of significance.

d) 90% two sided confidence interval on the difference:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(0.0091) - 1.65\sqrt{\frac{0.216(1-0.216)}{250} + \frac{0.207(1-0.207)}{290}} \le p_1 - p_2 \le (0.0091) + 1.65\sqrt{\frac{0.216(1-0.216)}{250} + \frac{0.207(1-0.207)}{290}}$$

$$-0.0491 \le p_1 - p_2 \le 0.0673$$

10-83 a) This is one-sided test because the hypotheses are $p_1 - p_2 = 0$ versus greater than 0.

b) $\hat{p}_1 = \frac{188}{250} = 0.752$ $\hat{p}_2 = \frac{245}{350} = 0.7$ $\hat{p} = \frac{188+245}{250+350} = 0.7217$

Test statistic is $z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$$z_0 = \frac{0.052}{\sqrt{(0.7217)(1-0.7217)\left(\frac{1}{250}+\frac{1}{350}\right)}} = 1.4012$$

P-value = [1- P(Z < 1.4012)] = 1 - 0.9194 = 0.0806

95% lower confidence interval on the difference:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2$$

$$(0.052) - 1.65\sqrt{\frac{0.752(1-0.752)}{250} + \frac{0.7(1-0.7)}{350}} \le p_1 - p_2$$

$$-0.0085 \le p_1 - p_2$$

c) The P-value = 0.0806 is less than α = 0.10. Therefore, we reject the null hypothesis that $p_1 - p_2 = 0$ at the 0.1 level of significance. If α = 0.05, the P-value = 0.0806 is greater than α = 0.05 and we fail to reject the null hypothesis.

10-84 a)

1) The parameters of interest are the proportion of successes of surgical repairs for different tears, p_1 and p_2

2) $H_0 : p_1 = p_2$

3) $H_1 : p_1 > p_2$

4) Test statistic is $$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

5) Reject the null hypothesis if $z_0 > z_{0.05}$ where $z_{0.05} = 1.65$ for $\alpha = 0.05$

6) $n_1 = 18$ $n_2 = 30$

$x_1 = 14$ $x_2 = 22$

$\hat{p}_1 = 0.78$ $\hat{p}_2 = 0.73$ $\hat{p} = \frac{14 + 22}{18 + 30} = 0.75$

$$z_0 = \frac{0.78 - 0.73}{\sqrt{0.75(1-0.75)\left(\frac{1}{18} + \frac{1}{30}\right)}} = 0.387$$

7) Conclusion: Because 0.387 < 1.65, we fail to reject the null hypothesis at the 0.05 level of significance.

P-value = $[1 - P(Z < 0.387)] = 1 - 0.6517 \approx 0.35$

b) 95% confidence interval on the difference:

$$(\hat{p}_1 - \hat{p}_2) - z_\alpha \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2$$

$$(0.78 - 0.73) - 1.65\sqrt{\frac{0.78(1-0.78)}{18} + \frac{0.73(1-0.73)}{30}} \le p_1 - p_2$$

$$-0.159 \le p_1 - p_2$$

Because this interval contains the value zero, there is not enough evidence to conclude that the success rate p_1 exceeds p_2.

10-85

The original version of this exercise is in error because the proportions are not independent. They come from the same sample. The test can be approximated based on a multinomial distribution. Let X_1, X_2, and X_3 denote the respondents that favor Bush, Kerry, and another candidate, respectively. If the population is large relative to the sample of 2020 respondents, one can assume that the joint distribution of X_1, X_2, X_3 is a multinomial distribution with parameters p_1, p_2, and p_3, respectively. The same size is n = 2020. The marginal distribution of both X_1 and X_2 is binomial with parameters p_1 and p_2, respectively, and sample size n =2020. However, X_1 and X_2 are not independent. Therefore, the sample proportions $\hat{p}_1 = \frac{X_1}{n}$ and $\hat{p}_2 = \frac{X_2}{n}$ are not independent. It can be shown that the covariance between X_1 and X_2 is $-np_1p_2$. Also, $V(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2) - 2Cov(\hat{p}_1, \hat{p}_2)$

Therefore,

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n} + \frac{p_1(1-p_1)}{n} + \frac{2p_1p_2}{n}$$

An approximate 95% confidence interval for the difference in proportions is

$$\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n} + \frac{2\hat{p}_1\hat{p}_2}{n}}$$

$$0.53 - 0.46 \pm 1.96\sqrt{\frac{0.53(1-0.53)}{2020} + \frac{0.46(1-0.46)}{2020} + \frac{2(0.53)(0.46)}{2020}}$$

= (–0.008, 0.148)

The revised exercise assumes that two samples are available with proportions 0.53 and 0.46 and both with the same sample size of 2020 (for 4040 respondents in total).

a)
1) The parameters of interest are the proportion of voters in favor of Bush and Kerry, p_1 and p_2, respectively
2) $H_0 : p_1 = p_2$
3) $H_1 : p_1 \neq p_2$
4) Test statistic is $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$
5) Reject the null hypothesis if $z_0 < -z_{0.025}$ or $z_0 > z_{0.025}$ where $z_{0.025} = 1.96$ for $\alpha = 0.05$
6) $n_1 = 2020$ $n_2 = 2020$
$x_1 = 1071$ $x_2 = 930$
$\hat{p}_1 = 0.53$ $\hat{p}_2 = 0.46$ $\hat{p} = \dfrac{1071 + 930}{2020 + 2020} = 0.495$

$$z_0 = \frac{0.53 - 0.46}{\sqrt{0.495(1-0.495)\left(\frac{1}{2020}+\frac{1}{2020}\right)}} = 4.45$$

7) Conclusion: Because 4.45 > 1.96, reject the null hypothesis and conclude that there is a difference in the proportions at the 0.05 level of significance.

P-value = $2[1 - P(Z < 4.45)] \approx 0$

b) 95% confidence interval on the difference

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$0.039 \le p_1 - p_2 \le 0.1$$

Because this interval does not contain the value zero, reject the hypothesis that the proportion are equal at the 0.05 level of significance.

10-86 a)
1) The parameters of interest are the proportion of defective parts, p_1 and p_2
2) $H_0 : p_1 = p_2$
3) $H_1 : p_1 \neq p_2$
4) Test statistic is $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$
5) Reject the null hypothesis if $z_0 < -z_{0.025}$ or $z_0 > z_{0.025}$ where $z_{0.025} = 1.96$ for $\alpha = 0.05$
6) $n_1 = 300$ $n_2 = 300$
$x_1 = 15$ $x_2 = 8$
$\hat{p}_1 = 0.05$ $\hat{p}_2 = 0.0267$ $\hat{p} = \dfrac{15 + 8}{300 + 300} = 0.0383$

$$z_0 = \frac{0.05 - 0.0267}{\sqrt{0.0383(1-0.0383)\left(\frac{1}{300}+\frac{1}{300}\right)}} = 1.49$$

7) Conclusion: Because −1.96 < 1.49 < 1.96, fail to reject the null hypothesis. There is not a significant difference in the fraction of defective parts produced by the two machines at the 0.05 level of significance.

P-value = 2[1 − P(*Z* < 1.49)] = 0.13622

b) 95% confidence interval on the difference:

$$(\hat{p}_1-\hat{p}_2)-z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\le p_1-p_2\le(\hat{p}_1-\hat{p}_2)+z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(0.05-0.0267)-1.96\sqrt{\frac{0.05(1-0.05)}{300}+\frac{0.0267(1-0.0267)}{300}}\le p_1-p_2\le(0.05-0.0267)+1.96\sqrt{\frac{0.05(1-0.05)}{300}+\frac{0.0267(1-0.0267)}{300}}$$

$$-0.0074\le p_1-p_2\le 0.054$$

Because this interval contains the value zero, there is no significant difference in the fraction of defective parts produced by the two machines. We have 95% confidence that the difference in proportions is between −0.0074 and 0.054.

c) Power = 1 − β

$$\beta=\Phi\left(\frac{z_{\alpha/2}\sqrt{\overline{p}\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}}\right)-\Phi\left(\frac{-z_{\alpha/2}\sqrt{\overline{p}\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}}\right)$$

$$\overline{p}=\frac{300(0.05)+300(0.01)}{300+300}=0.03 \qquad \overline{q}=0.97$$

$$\hat{\sigma}_{\hat{p}_1-\hat{p}_2}=\sqrt{\frac{0.05(1-0.05)}{300}+\frac{0.01(1-0.01)}{300}}=0.014$$

$$\beta=\Phi\left(\frac{1.96\sqrt{0.03(0.97)\left(\frac{1}{300}+\frac{1}{300}\right)}-(0.05-0.01)}{0.014}\right)-\Phi\left(\frac{-1.96\sqrt{0.03(0.97)\left(\frac{1}{300}+\frac{1}{300}\right)}-(0.05-0.01)}{0.014}\right)$$

$$=\Phi(-0.91)-\Phi(-4.81)=0.18141-0=0.18141$$

Power = 1 − 0.18141 = 0.81859

d)
$$n=\frac{\left(z_{\alpha/2}\sqrt{\frac{(p_1+p_2)(q_1+q_2)}{2}}+z_\beta\sqrt{p_1q_1+p_2q_2}\right)^2}{(p_1-p_2)^2}$$

$$=\frac{\left(1.96\sqrt{\frac{(0.05+0.01)(0.95+0.99)}{2}}+1.29\sqrt{0.05(0.95)+0.01(0.99)}\right)^2}{(0.05-0.01)^2}=382.11$$

n = 383

e)
$$\beta=\Phi\left(\frac{z_{\alpha/2}\sqrt{\overline{p}\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}}\right)-\Phi\left(\frac{-z_{\alpha/2}\sqrt{\overline{p}\overline{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}-(p_1-p_2)}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}}\right)$$

$$\overline{p}=\frac{300(0.05)+300(0.02)}{300+300}=0.035 \qquad \overline{q}=0.965$$

$$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.05(1-0.05)}{300}+\frac{0.02(1-0.02)}{300}} = 0.015$$

$$\beta = \Phi\left(\frac{1.96\sqrt{0.035(0.965)\left(\frac{1}{300}+\frac{1}{300}\right)}-(0.05-0.02)}{0.015}\right)-\Phi\left(\frac{-1.96\sqrt{0.035(0.965)\left(\frac{1}{300}+\frac{1}{300}\right)}-(0.05-0.02)}{0.015}\right)$$

$= \Phi(-0.04)-\Phi(-3.96) = 0.48405-0.00004 = 0.48401$

Power = 1 – 0.48401 = 0.51599

f) $$n = \frac{\left(z_{\alpha/2}\sqrt{\frac{(p_1+p_2)(q_1+q_2)}{2}}+z_\beta\sqrt{p_1q_1+p_2q_2}\right)^2}{(p_1-p_2)^2}$$

$$= \frac{\left(1.96\sqrt{\frac{(0.05+0.02)(0.95+0.98)}{2}}+1.29\sqrt{0.05(0.95)+0.02(0.98)}\right)^2}{(0.05-0.02)^2} = 790.67$$

n = 791

10-87 a)

1) The parameters of interest are the proportion of satisfactory lenses, p_1 and p_2

2) $H_0 : p_1 = p_2$

3) $H_1 : p_1 \neq p_2$

4) Test statistic is

$$z_0 = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad \text{where} \quad \hat{p} = \frac{x_1+x_2}{n_1+n_2}$$

5) Reject the null hypothesis if $z_0 < -z_{0.005}$ or $z_0 > z_{0.005}$ where $z_{0.005} = 2.58$ for $\alpha = 0.01$

6) $n_1 = 300$ $n_2 = 300$

$x_1 = 253$ $x_2 = 196$

$\hat{p}_1 = 0.843$ $\hat{p}_2 = 0.653$ $\hat{p} = \frac{253+196}{300+300} = 0.748$

$$z_0 = \frac{0.843-0.653}{\sqrt{0.748(1-0.748)\left(\frac{1}{300}+\frac{1}{300}\right)}} = 5.36$$

7) Conclusion: Because 5.36 > 2.58, reject the null hypothesis and conclude that there is a difference in the fraction of polishing-induced defects produced by the two polishing solutions at the 0.01 level of significance.

P-value = 2[1 – P(*Z* < 5.36)] ≈ 0

b) By constructing a 99% confidence interval on the difference in proportions, the same question can be answered by whether or not zero is contained in the interval.

10-88 a)

1) The parameters of interest are the proportion of residents in favor of an increase, p_1 and p_2

2) $H_0 : p_1 = p_2$

3) $H_1 : p_1 \neq p_2$

4) Test statistic is $$z_0 = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad \text{where} \quad \hat{p} = \frac{x_1+x_2}{n_1+n_2}$$

5) Reject the null hypothesis if $z_0 < -z_{0.025}$ or $z_0 > z_{0.025}$ where $z_{0.025} = 1.96$ for $\alpha = 0.05$

6) $n_1 = 500$ $n_2 = 400$

$x_1 = 385$ $x_2 = 267$

$\hat{p}_1 = 0.77$ $\hat{p}_2 = 0.6675$ $\hat{p} = \dfrac{385+267}{500+400} = 0.724$

$$z_0 = \frac{0.77 - 0.6675}{\sqrt{0.724(1-0.724)\left(\frac{1}{500}+\frac{1}{400}\right)}} = 3.42$$

7) Conclusion: Because 3.42 > 1.96, reject the null hypothesis and conclude that there is a difference in the proportions of support for increasing the speed limit between residents of the two counties at the 0.05 level of significance.

P-value = 2[1 – P(*Z* < 3.42)] = 0.00062

b) 95% confidence interval on the difference:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(0.77-0.6675) - 1.96\sqrt{\frac{0.77(1-0.77)}{500}+\frac{0.6675(1-0.6675)}{400}} \le p_1 - p_2 \le (0.77-0.6675) + 1.96\sqrt{\frac{0.77(1-0.77)}{500}+\frac{0.6675(1-0.6675)}{400}}$$

$$0.0434 \le p_1 - p_2 \le 0.1616$$

We are 95% confident that the difference in proportions is between 0.0434 and 0.1616. Because the interval does not contain zero there is evidence that the counties differ in support of the change.

10-89

$H_0: p_1 = p_2$ where “1” refers to WTC and “2” refers to the other hospital

$H_1: p_1 > p_2$

$\hat{p}_1 = \dfrac{15}{182} = 0.082$ $\hat{p}_2 = \dfrac{92}{2300} = 0.04$ $\hat{p} = \dfrac{15+92}{182+2300} = 0.043$

Test statistic is $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where, $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

$$z_0 = \frac{0.082 - 0.04}{\sqrt{(0.43)(1-0.043)\left(\frac{1}{182}+\frac{1}{2300}\right)}} = 2.712$$

P-value = [1 - Φ(2.712)] = 0.003

Because P-value < α = 0.05, reject H_0. There is sufficient evidence to conclude that the exposed mothers had a higher incidence of low-weight babies at α = 0.05.

10-90

$H_0: p_1 = p_2$

$H_1: p_1 < p_2$

$\hat{p}_1 = \dfrac{18}{347} = 0.052$ $\hat{p}_2 = \dfrac{31}{348} = 0.089$ $\hat{p} = \dfrac{18+31}{347+348} = 0.071$

Test statistic is $z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where, $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

$$z_0 = \frac{0.052 - 0.089}{\sqrt{(0.071)(1-0.071)\left(\frac{1}{347} + \frac{1}{348}\right)}} = -1.916$$

P-value = Φ(-1.916) = 0.028.

Because P-value < α = 0.05, reject H_0. There is sufficient evidence to conclude that the surgery decreased the proportion of those who died of prostate cancer at α = 0.05.

10-91

$\tilde{n}_1 =$ 2020+2=2022 $\quad \tilde{n}_2 =$ 2020+2=2022

$x_1 + 1 =$ 1071+1=1072 $\quad x_2 + 1 =$ 930+1=931

$$\tilde{p}_1 = \frac{x_1+1}{n_1+1} = \frac{1072}{2021} = 0.53 \qquad \tilde{p}_2 = \frac{x_2+1}{n_2+1} = \frac{931}{2021} = 0.461$$

95% confidence interval on the difference

$$(\tilde{p}_1 - \tilde{p}_2) - z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}} \le p_1 - p_2 \le (\tilde{p}_1 - \tilde{p}_2) + z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}}$$

$0.0383 \le p_1 - p_2 \le 0.0997$

This interval is similar to the previous interval which was $0.039 \le p_1 - p_2 \le 0.1$. The coverage level of the new interval is expected to be closer to the advertised level of 95%.

10-92

$\tilde{n}_1 =$ 500+2=502 $\quad \tilde{n}_2 =$ 400+2=402

$x_1 + 1 =$ 385+1=386 $\quad x_2 + 1 =$ 267+1=268

$$\tilde{p}_1 = \frac{x_1+1}{n_1+1} = \frac{386}{501} = 0.77 \qquad \tilde{p}_2 = \frac{x_2+1}{n_2+1} = \frac{268}{401} = 0.668$$

99% confidence interval on the difference:

$$(\tilde{p}_1 - \tilde{p}_2) - z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}} \le p_1 - p_2 \le (\tilde{p}_1 - \tilde{p}_2) + z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}}$$

$0.0244 \le p_1 - p_2 \le 0.1795$

This is wider than the previous interval which was $0.0434 \le p_1 - p_2 \le 0.1616$. Because the previous interval is constructed for 95% confidence, it is expected that the new interval is wider so that its coverage level is closer to the advertised level of 99%.

Supplemental Exercises

10-93 a) SE Mean$_1$ $= \frac{s_1}{\sqrt{n_1}} = \frac{2.23}{\sqrt{20}} = 0.50$

$\bar{x}_1 = 11.87 \quad \bar{x}_2 = 12.73 \quad s_1^2 = 2.23^2 \quad s_2^2 = 3.19^2 \quad n_1 = 20 \quad n_2 = 20$

Degrees of freedom = $n_1 + n_2$ - 2 = 20 – 20 - 2 = 38.

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(20-1)2.23^2 + (20-1)3.19^2}{20+20-2}} = 2.7522$$

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(-0.86)}{2.7522\sqrt{\frac{1}{20} + \frac{1}{20}}} = -0.9881$$

P-value = 2 [P(t < –0.9881)] and 2(0.10) <*P*-value < 2(0.25) = 0.20 < *P*-value < 0.5

The 95% two-sided confidence interval: $t_{\alpha/2,n_1+n_2-2} = t_{0.025,38} = 2.024$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(-0.86) - (2.024)(2.7522)\sqrt{\frac{1}{20}+\frac{1}{20}} \le \mu_1 - \mu_2 \le (-0.86) + (2.024)(2.7522)\sqrt{\frac{1}{20}+\frac{1}{20}}$$

$$-2.622 \le \mu_1 - \mu_2 \le 0.902$$

b) This is two-sided test because the alternative hypothesis is $\mu_1 - \mu_2$ not = 0.

c) Because the 0.20 < *P*-value < 0.5 and the P-value > α = 0.05, we fail to reject the null hypothesis at the 0.05 level of significance. If α = 0.01, we also fail to reject the null hypothesis.

10-94 a) This is one-sided test because the alternative hypothesis is $\mu_1 - \mu_2 < 0$.

b)

$$\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}} = \frac{\left(\frac{2.98^2}{16}+\frac{5.36^2}{25}\right)^2}{\frac{(2.98^2/16)^2}{16-1}+\frac{(5.36^2/25)^2}{25-1}} = 38.44 \approx 38 \text{ (truncated)}$$

Degrees of freedom = 38
P-value = P(t < –1.65) and 0.05 < *P*-value < 0.1

c) Because 0.05 < *P*-value < 0.1 and the P-value > α = 0.05, we fail to reject the null hypothesis of $\mu_1 - \mu_2 = 0$ at the 0.05 level of significance. If α = 0.1, we reject the null hypothesis because the *P*-value < 0.1.

d) The 95% upper one-sided confidence interval: $t_{0.05,38} = 1.686$

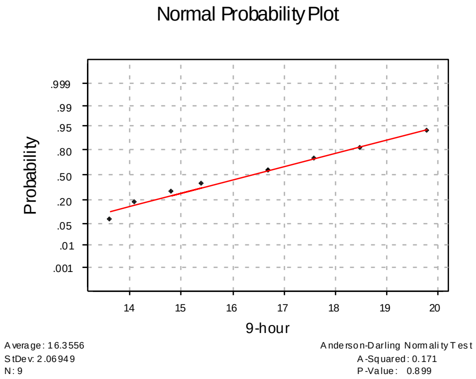
$$\mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

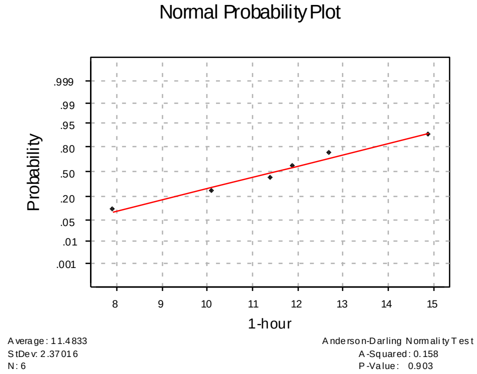
$$\mu_1 - \mu_2 \le (-2.16) + 1.686\sqrt{\frac{(2.98)^2}{16}+\frac{(5.36)^2}{25}}$$

$$\mu_1 - \mu_2 \le 0.0410$$

10-95 a) Assumptions that must be met are normality, equality of variance, and independence of the observations. Normality and equality of variances appear to be reasonable from the normal probability plots. The data appear to fall along lines and the slopes appear to be the same. Independence of the observations for each sample is obtained if random samples are selected.

.





b) $\bar{x}_1 = 16.36 \qquad \bar{x}_2 = 11.483$

$s_1 = 2.07 \qquad s_2 = 2.37$

$n_1 = 9 \quad n_2 = 6$

99% confidence interval: $t_{\alpha/2,n_1+n_2-2} = t_{0.005,13}$ where $t_{0.005,13} = 3.012$

$$s_p = \sqrt{\frac{8(2.07)^2 + 5(2.37)^2}{13}} = 2.19$$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,n_1+n_2-2}(s_p)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(16.36 - 11.483) - 3.012(2.19)\sqrt{\frac{1}{9}+\frac{1}{6}} \le \mu_1 - \mu_2 \le (16.36 - 11.483) + 3.012(2.19)\sqrt{\frac{1}{9}+\frac{1}{6}}$$

$$1.40 \le \mu_1 - \mu_2 \le 8.36$$

c) Yes, we are 99% confident the results from the first test condition exceed the results of the second test condition because the confidence interval contains only positive values.

d) 95% confidence interval for σ_1^2 / σ_2^2

$$f_{0.975,8,5} = \frac{1}{f_{0.025,5,8}} = \frac{1}{4.82} = 0.2075, \qquad f_{0.025,8,5} = 6.76$$

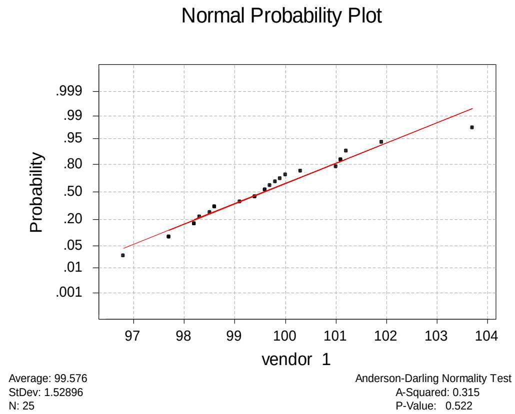
$$\frac{s_1^2}{s_2^2} f_{0.975,5,8} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} f_{0.025,5,8}$$

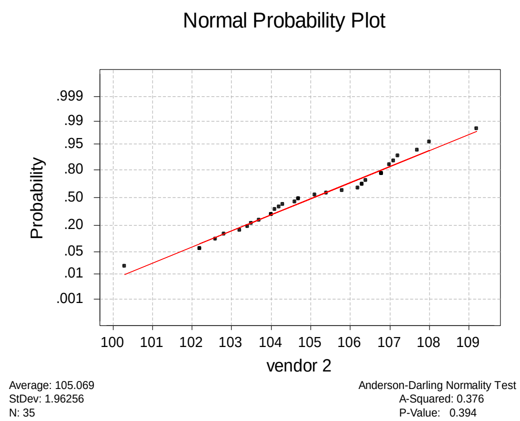
$$\left(\frac{4.283}{5.617}\right)(0.148) \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{4.283}{5.617}\right)(4.817)$$

$$0.113 \le \frac{\sigma_1^2}{\sigma_2^2} \le 3.673$$

e) Because the value one is contained within this interval, the population variances do not differ at a 5% significance level.

10-96 a) Assumptions that must be met are normality and independence of the observations. Normality appears to be reasonable.





The data appear to fall along lines in the normal probability plots. Because the slopes appear to be the same, it appears the population standard deviations are similar. Independence of the observations for each sample is obtained if random samples are selected.

b)

1) The parameters of interest are the variances of resistance of products, σ_1^2, σ_2^2

2) $H_0 : \sigma_1^2 = \sigma_2^2$

3) $H_1 : \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject H_0 if $f_0 < f_{0.975,24,34}$ where $f_{0.975,24,34} = \dfrac{1}{f_{0.025,34,24}} = \dfrac{1}{2.18} = 0.459$ for $\alpha = 0.05$

or $f_0 > f_{0.025,24,34}$ where $f_{0.025,24,34} = 2.07$ for $\alpha = 0.05$

6) $s_1 = 1.53$ $s_2 = 1.96$
$n_1 = 25$ $n_2 = 35$

$$f_0 = \frac{(1.53)^2}{(1.96)^2} = 0.609$$

7) Conclusion: Because $0.459 < 0.609 < 2.07$, fail to reject H_0. There is not sufficient evidence to conclude that the variances are different at $\alpha = 0.05$.

10-97 a)

1) The parameter of interest is the mean weight loss, μ_d, where d_i = Initial Weight – Final Weight.

2) $H_0 : \mu_d = 3$
3) $H_1 : \mu_d > 3$
4) The test statistic is

$$t_0 = \frac{\bar{d} - \Delta_0}{s_d / \sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,7} = 1.895$for $\alpha = 0.05$.
6) $\bar{d} = 4.125$
$s_d = 1.246$
$n = 8$

$$t_0 = \frac{4.125 - 3}{1.246 / \sqrt{8}} = 2.554$$

7) Conclusion: Because 2.554 > 1.895, reject the null hypothesis and conclude the mean weight loss is greater than 3 at $\alpha = 0.05$.

b)
2) $H_0 : \mu_d = 3$
3) $H_1 : \mu_d > 3$
4) The test statistic is

$$t_0 = \frac{\bar{d} - \Delta_0}{s_d / \sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.01,7} = 2.998$ for $\alpha = 0.01$.
6) $\bar{d} = 4.125$
$s_d = 1.246$
$n = 8$

$$t_0 = \frac{4.125 - 3}{1.246 / \sqrt{8}} = 2.554$$

7) Conclusion: Because 2.554 < 2.998, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the mean weight loss is greater than 3 at $\alpha = 0.01$.

c)
2) $H_0 : \mu_d = 5$
3) $H_1 : \mu_d > 5$
4) The test statistic is

$$t_0 = \frac{\bar{d} - \Delta_0}{s_d / \sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.05,7} =1.895$ for $\alpha = 0.05$
6) $\bar{d} = 4.125$
$s_d = 1.246$
$n = 8$

$$t_0 = \frac{4.125 - 5}{1.246 / \sqrt{8}} = -1.986$$

7) Conclusion: Because −1.986 < 1.895, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the mean weight loss is greater than 5 at $\alpha = 0.05$.

d)
2) $H_0 : \mu_d = 5$
3) $H_1 : \mu_d > 5$
4) The test statistic is

$$t_0 = \frac{\bar{d} - \Delta_0}{s_d / \sqrt{n}}$$

5) Reject H_0 if $t_0 > t_{\alpha,n-1}$ where $t_{0.01,7} = 2.998$ for $\alpha = 0.01$.

6) $\bar{d} = 4.125$
$s_d = 1.246$
$n = 8$

$$t_0 = \frac{4.125 - 5}{1.246/\sqrt{8}} = -1.986$$

7) Conclusion: Because –1.986 < 2.998, fail to reject the null hypothesis. There is not sufficient evidence to conclude that the mean weight loss is greater than 5 at $\alpha = 0.01$.

10-98 $$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

a) 90% confidence interval: $z_{\alpha/2} = 1.65$

$$(88-91) - 1.65\sqrt{\frac{5^2}{20} + \frac{4^2}{20}} \le \mu_1 - \mu_2 \le (88-91) + 1.65\sqrt{\frac{5^2}{20} + \frac{4^2}{20}}$$
$$-5.362 \le \mu_1 - \mu_2 \le -0.638$$

Yes, the data indicate that the mean breaking strength of the yarn of manufacturer 2 exceeds that of manufacturer 1 by between 5.362 and 0.638 with 90% confidence.

b) 98% confidence interval: $z_{\alpha/2} = 2.33$

$$(88-91) - 2.33\sqrt{\frac{5^2}{20} + \frac{4^2}{20}} \le \mu_1 - \mu_2 \le (88-91) + 2.33\sqrt{\frac{5^2}{20} + \frac{4^2}{20}}$$
$$-6.340 \le \mu_1 - \mu_2 \le 0.340$$

Because the confidence interval contains zero, we cannot conclude that one yarn has greater mean breaking strength at significance level 0.02.

c) The results of parts (a) and (b) are different because the confidence level used is different. The appropriate interval depends upon the level of confidence considered acceptable.

10-99 a)
1) The parameters of interest are the proportions of children who contract polio, p_1 , p_2
2) $H_0 : p_1 = p_2$
3) $H_1 : p_1 \neq p_2$
4) The test statistic is

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$ where $z_{\alpha/2} = 1.96$ for $\alpha = 0.05$

6) $\hat{p}_1 = \frac{x_1}{n_1} = \frac{110}{201299} = 0.00055$ (Placebo) $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = 0.000356$

$\hat{p}_2 = \frac{x_2}{n_2} = \frac{33}{200745} = 0.00016$ (Vaccine)

$$z_0 = \frac{0.00055 - 0.00016}{\sqrt{0.000356(1-0.000356)\left(\frac{1}{201299} + \frac{1}{200745}\right)}} = 6.55$$

7) Because 6.55 > 1.96, reject H_0 and conclude the proportions of children who contracted polio differ at $\alpha = 0.05$.

b) $\alpha = 0.01$
Reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$ where $z_{\alpha/2}$ =2.58. Here, still $z_0 = 6.55$.
Because 6.55 > 2.58, reject H_0 and conclude the proportions of children who contracted polio differ at $\alpha = 0.01$.

c) The conclusions are the same because z_0 is large enough to exceed $z_{\alpha/2}$ in both cases.

10-100 a) $\alpha = 0.10$ $z_{\alpha/2} = 1.65$

$$n \cong \frac{(z_{\alpha/2})^2(\sigma_1^2+\sigma_2^2)}{(E)^2} \cong \frac{(1.65)^2(25+16)}{(1.5)^2} = 49.61, \quad n = 50$$

b) $\alpha = 0.02$ $z_{\alpha/2} = 2.33$

$$n \cong \frac{(z_{\alpha/2})^2(\sigma_1^2+\sigma_2^2)}{(E)^2} \cong \frac{(2.33)^2(25+16)}{(1.5)^2} = 98.93, \quad n = 99$$

c) As the confidence level increases, the sample size also increases.

d) $\alpha = 0.10$ $z_{\alpha/2} = 1.65$

$$n \cong \frac{(z_{\alpha/2})^2(\sigma_1^2+\sigma_2^2)}{(E)^2} \cong \frac{(1.65)^2(25+16)}{(0.75)^2} = 198.44, \quad n = 199$$

$\alpha = 0.02$ $z_{\alpha/2} = 2.33$

$$n \cong \frac{(z_{\alpha/2})^2(\sigma_1^2+\sigma_2^2)}{(E)^2} \cong \frac{(2.33)^2(25+16)}{(0.75)^2} = 395.70, \quad n = 396$$

e) As the error decreases, the required sample size increases.

10-101 $\hat{p}_1 = \frac{x_1}{n_1} = \frac{387}{1500} = 0.258$ $\hat{p}_2 = \frac{x_2}{n_2} = \frac{310}{1200} = 0.2583$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

a) $z_{\alpha/2} = z_{0.025} = 1.96$

$$(0.258 - 0.2583) \pm 1.96\sqrt{\frac{0.258(0.742)}{1500} + \frac{0.2583(0.7417)}{1200}}$$

$$-0.0335 \le p_1 - p_2 \le 0.0329$$

Because zero is contained in this interval, there is no significant difference between the proportions of unlisted numbers in the two cities at a 5% significance level.

b) $z_{\alpha/2} = z_{0.05} = 1.65$

$$(0.258 - 0.2583) \pm 1.65\sqrt{\frac{0.258(0.742)}{1500} + \frac{0.2583(0.7417)}{1200}}$$

$$-0.0282 \le p_1 - p_2 \le 0.0276$$

The proportions of unlisted numbers in the two cities do not significantly differ at a 5% significance level.

c) $\hat{p}_1 = \frac{x_1}{n_1} = \frac{774}{3000} = 0.258$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{620}{2400} = 0.2583$$

95% confidence interval:

$$(0.258 - 0.2583) \pm 1.96\sqrt{\frac{0.258(0.742)}{3000} + \frac{0.2583(0.7417)}{2400}}$$

$$-0.0238 \le p_1 - p_2 \le 0.0232$$

90% confidence interval:

$$\left(0.258-0.2583\right)\pm 1.65\sqrt{\frac{0.258(0.742)}{3000}+\frac{0.2583(0.7417)}{2400}} \qquad -0.0201\le p_1-p_2\le 0.0195$$

Increasing the sample size decreased the width of the confidence interval, but did not change the conclusions drawn. The conclusion remains that there is no significant difference.

10-102 a)

1) The parameters of interest are the proportions of those residents who wear a seat belt regularly, p_1 , p_2
2) $H_0 : p_1 = p_2$
3) $H_1 : p_1 \neq p_2$
4) The test statistic is

$$z_0=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$ where $z_{0.025} = 1.96$ for $\alpha = 0.05$

6) $\hat{p}_1=\frac{x_1}{n_1}=\frac{165}{200}=0.825$ $\qquad \hat{p}=\frac{x_1+x_2}{n_1+n_2}=0.807$

$$\hat{p}_2=\frac{x_2}{n_2}=\frac{198}{250}=0.792$$

$$z_0=\frac{0.825-0.792}{\sqrt{0.807(1-0.807)\left(\frac{1}{200}+\frac{1}{250}\right)}}=0.8814$$

7) Conclusion: Because $-1.96 < 0.8814 < 1.96$, fail to reject H_0. There is not sufficient evidence that there is a difference in seat belt usage at $\alpha = 0.05$.

b) $\alpha = 0.10$
Reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$ where $z_{0.05} = 1.65$ $\qquad z_0 = 0.8814$

Because $-1.65 < 0.8814 < 1.65$, fail to reject H_0. There is not sufficient evidence that there is a difference in seat belt usage at $\alpha = 0.10$.

c) The conclusions are the same, but with different levels of significance.

d) $n_1 = 400$, $n_2 = 500$

$\alpha = 0.05$
Reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$ where $z_{0.025} = 1.96$

$$z_0=\frac{0.825-0.792}{\sqrt{0.807(1-0.807)\left(\frac{1}{400}+\frac{1}{500}\right)}}=1.246$$

Because $-1.96 < 1.246 < 1.96$, fail to reject H_0. There is not sufficient evidence that there is a difference in seat belt usage at $\alpha = 0.05$.

$\alpha = 0.10$
Reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$ where $z_{0.05} = 1.65$ $\qquad z_0 = 1.246$
Because $-1.65 < 1.246 < 1.65$, fail to reject H_0. There is not sufficient evidence that there is a difference in seat belt usage at $\alpha = 0.10$.

As the sample size increased, the test statistic also increased (because the denominator of z_0 decreased). However, the sample size increase was not enough to change our conclusion.

10-103 a) Yes, there could be some bias in the results due to the telephone survey.
b) If it could be shown that these populations are similar to the respondents, the results may be extended.

10-104 The parameter of interest is $\mu_1 - 2\mu_2$

$$H_0: \mu_1 = 2\mu_2 \quad \rightarrow \quad H_0: \mu_1 - 2\mu_2 = 0$$
$$H_1: \mu_1 > 2\mu_2 \qquad\quad H_1: \mu_1 - 2\mu_2 > 0$$

Let n_1 = size of sample 1 $\qquad \overline{X}_1$ estimate for μ_1

Let n_2 = size of sample 2 $\qquad \overline{X}_2$ estimate for μ_2

$\overline{X}_1 - 2\overline{X}_2$ is an estimate for $\mu_1 - 2\mu_2$

The variance is V($\overline{X}_1 - 2\overline{X}_2$) = V($\overline{X}_1$) + V(2 $\overline{X}_2$) = $\frac{\sigma_1^2}{n_1} + \frac{4\sigma_2^2}{n_2}$

The test statistic for this hypothesis is:

$$Z_0 = \frac{(\overline{X}_1 - 2\overline{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{4\sigma_2^2}{n_2}}}$$

We reject the null hypothesis if $z_0 > z_{\alpha/2}$ for a given level of significance. *P*-value = $P(Z \ge z_0)$.

10-105 $\bar{x}_1 = 30.87 \quad \bar{x}_2 = 30.68$

$\sigma_1 = 0.10 \quad \sigma_2 = 0.15$

$n_1 = 12 \quad n_2 = 10$

a) 90% two-sided confidence interval:

$$\left(\bar{x}_1 - \bar{x}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(30.87 - 30.68) - 1.645\sqrt{\frac{(0.10)^2}{12} + \frac{(0.15)^2}{10}} \le \mu_1 - \mu_2 \le \left(30.87 - 30.68\right) + 1.645\sqrt{\frac{(0.10)^2}{12} + \frac{(0.15)^2}{10}}$$

$$0.0987 \le \mu_1 - \mu_2 \le 0.2813$$

We are 90% confident that the mean fill volume for machine 1 exceeds that of machine 2 by between 0.0987 and 0.2813 fl. oz.

b) 95% two-sided confidence interval:

$$\left(\bar{x}_1 - \bar{x}_2\right) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(30.87 - 30.68) - 1.96\sqrt{\frac{(0.10)^2}{12} + \frac{(0.15)^2}{10}} \le \mu_1 - \mu_2 \le \left(30.87 - 30.68\right) + 1.96\sqrt{\frac{(0.10)^2}{12} + \frac{(0.15)^2}{10}}$$

$$0.0812 \le \mu_1 - \mu_2 \le 0.299$$

We are 95% confident that the mean fill volume for machine 1 exceeds that of machine 2 by between 0.0812 and 0.299 fl. oz.

Comparison of parts (a) and (b): As the level of confidence increases, the interval width also increases (with all other variables held constant).

c) 95% upper-sided confidence interval:

$$\mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + z_{\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\mu_1 - \mu_2 \le \left(30.87 - 30.68\right) + 1.645\sqrt{\frac{(0.10)^2}{12} + \frac{(0.15)^2}{10}}$$

$$\mu_1 - \mu_2 \le 0.2813$$

With 95% confidence, the fill volume for machine 1 exceeds the fill volume of machine 2 by no more than 0.2813 fl. oz.

d)
1) The parameter of interest is the difference in mean fill volume $\mu_1 - \mu_2$
2) H_0 : $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) H_1 : $\mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$
4) The test statistic is

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2} = -1.96$ or $z_0 > z_{\alpha/2} = 1.96$ for $\alpha = 0.05$
6) $\bar{x}_1 = 30.87$ $\bar{x}_2 = 30.68$
$\sigma_1 = 0.10$ $\sigma_2 = 0.15$
$n_1 = 12$ $n_2 = 10$

$$z_0 = \frac{(30.87 - 30.68)}{\sqrt{\frac{(0.10)^2}{12} + \frac{(0.15)^2}{10}}} = 3.42$$

7) Because 3.42 > 1.96, reject the null hypothesis and conclude the mean fill volumes of machine 1 and machine 2 differ at $\alpha = 0.05$.

$$P\text{-value} = 2[1 - \Phi(3.42)] = 2(1 - 0.99969) = 0.00062$$

e) Assume the sample sizes are to be equal, use $\alpha = 0.05$, $\beta = 0.10$, and $\Delta = 0.20$

$$n \cong \frac{\left(z_{\alpha/2} + z_\beta\right)^2\left(\sigma_1^2 + \sigma_2^2\right)}{\left(\Delta - \Delta_0\right)^2} = \frac{(1.96 + 1.28)^2\left((0.10)^2 + (0.15)^2\right)}{(-0.20)^2} = 8.53, \quad n = 9, \text{ use } n_1 = n_2 = 9$$

10-106 $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$
$n_1 = n_2 = n$
$\beta = 0.10$
$\alpha = 0.05$
Assume normal distribution and $\sigma_1^2 = \sigma_2^2 = \sigma^2$
$\mu_1 = \mu_2 + \sigma$

$$d = \frac{|\mu_1 - \mu_2|}{2\sigma} = \frac{\sigma}{2\sigma} = \frac{1}{2}$$

From Chart VII*e*, $n^* = 50$ and $n = \frac{n^* + 1}{2} = \frac{50 + 1}{2} = 25.5$ and $n_1 = n_2 = 26$

10-107 a)
1) The parameters of interest are: the proportion of lenses that are unsatisfactory after tumble-polishing, p_1, p_2
2) $H_0 : p_1 = p_2$
3) $H_1 : p_1 \neq p_2$
4) The test statistic is

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

5) Reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$ where $z_{\alpha/2} = 2.58$ for $\alpha = 0.01$.
6) x_1 = number of defective lenses

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{47}{300} = 0.1567 \qquad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = 0.2517$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{104}{300} = 0..3467$$

$$z_0 = \frac{0.1567 - 0.3467}{\sqrt{0.2517(1-0.2517)\left(\frac{1}{300}+\frac{1}{300}\right)}} = -5.36$$

7) Conclusion: Because −5.36 < −2.58, reject H_0 and conclude that the proportions from the two polishing fluids are different at $\alpha = 0.01$.

b) The conclusions are the same whether we analyze the data using the proportion unsatisfactory or proportion satisfactory.

c)

$$n = \frac{\left(2.575\sqrt{\frac{(0.9+0.6)(0.1+0.4)}{2}} + 1.28\sqrt{0.9(0.1)+0.6(0.4)}\right)^2}{(0.9-0.6)^2}$$

$$= \frac{5.346}{0.09} = 59.4$$

$$n = 60$$

10-108 a) $\alpha = 0.05$, $\beta = 0.05$, $\Delta = 1.5$. Use $s_p = 0.7071$ to approximate σ.

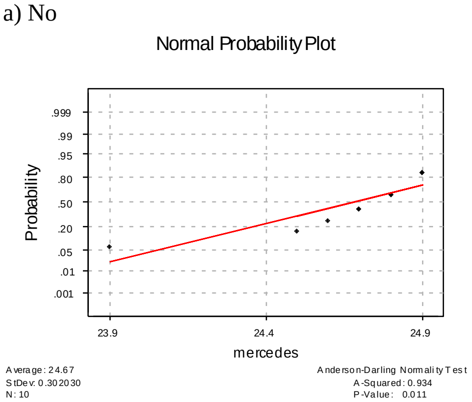
$$d = \frac{\Delta}{2(s_p)} = \frac{1.5}{2(.7071)} = 1.06 \cong 1$$

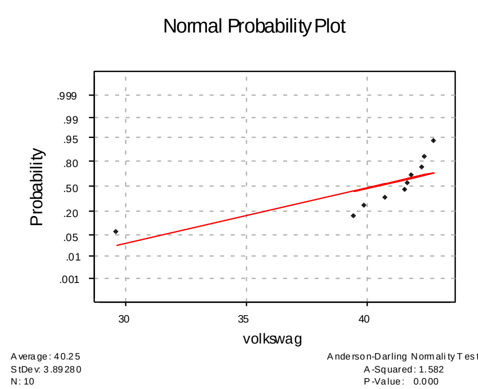
From Chart VII*e*, $n^* = 20$ $n = \frac{n^*+1}{2} = \frac{20+1}{2} = 10.5$ n = 11

is needed to detect that the two agents differ by 0.5 with probability of at least 0.95.

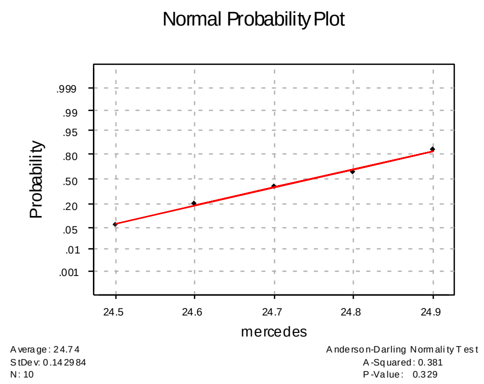
b) The original size of n = 5 was not appropriate to detect the difference because a sample size of 11 is needed to detect that the two agents differ by 1.5 with probability of at least 0.95.

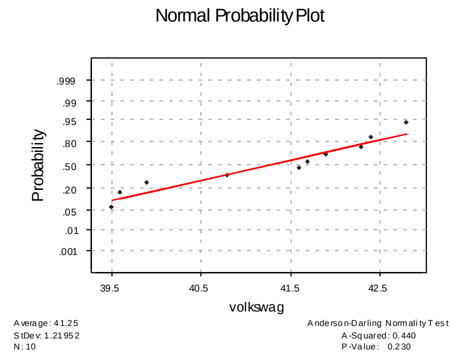
10-109 a) No





b) Normal distributions are reasonable because the data appear to fall along lines on the normal probability plots. The plots also indicate that the variances appear to be equal because the slopes appear to be the same.





c) By correcting the data points, it is more apparent the data follow normal distributions. Note that one unusual observation can cause an analyst to reject the normality assumption.

d) Consider a one-sided 95% confidence interval on the ratio of the variances, σ_V^2 / σ_M^2

$$s_V^2 = 1.49$$

$$s_M^2 = 0.0204 \qquad f_{9,9,0.95} = \frac{1}{f_{9,9,0.05}} = \frac{1}{3.18} = 0.314$$

$$\left(\frac{s_V^2}{s_M^2}\right) f_{9,9,0.95} < \frac{\sigma_V^2}{\sigma_M^2}$$

$$\left(\frac{1.49}{0.0204}\right) 0.314 < \frac{\sigma_V^2}{\sigma_M^2}$$

$$22.93 < \frac{\sigma_V^2}{\sigma_M^2}$$

Because the interval does not include the value one, we reject the hypothesis that variability in mileage performance is the same for the two types of vehicles. There is evidence that the variability is greater for a Volkswagen than for a Mercedes.

e)

1) The parameters of interest are the variances in mileage performance, σ_V^2, σ_M^2

2) $H_0 : \sigma_V^2 = \sigma_M^2$

3) $H_1 : \sigma_V^2 > \sigma_M^2$

4) The test statistic is

$$f_0 = \frac{s_V^2}{s_M^2}$$

5) Reject H_0 if $f_0 > f_{0.05,9,9}$ where $f_{0.05,9,9} = 3.18$ for $\alpha = 0.05$

6) $s_1 = 1.22$ $\quad$ $s_2 = 0.143$
$n_1 = 10$ $\quad$ $n_2 = 10$

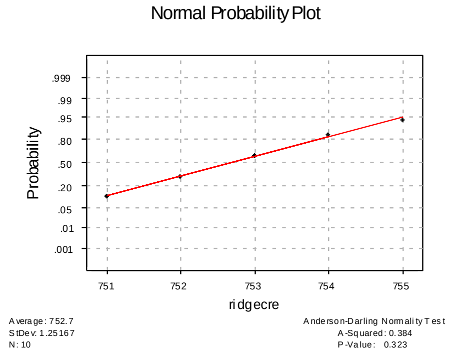
$$f_0 = \frac{(122)^2}{(0.143)^2} = 72.78$$

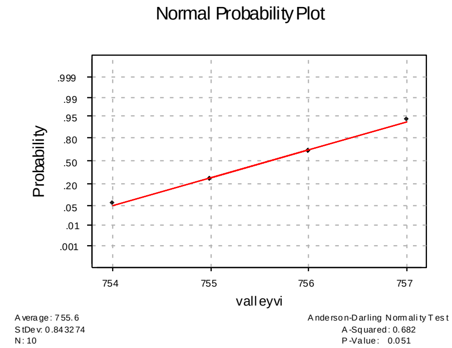
7) Conclusion: Because 72.78 > 3.28, reject H_0 and conclude that there is a difference between Volkswagen and Mercedes in terms of mileage variability. The same conclusions are reached in part (d).

Recall $f_{9,9,0.05} = \dfrac{1}{f_{9,9,0.95}}$. The hypothesis test rejects when $f_0 = \dfrac{s_V^2}{s_M^2} > f_{0.05,9,9}$ and this is equivalent to $\dfrac{s_V^2}{s_M^2} > \dfrac{1}{f_{0.95,9,9}}$ and this is also equivalent to the lower confidence limit $\dfrac{s_V^2}{s_M^2} f_{0.95,9,9} > 1$. Consequently the tests rejects when the lower confidence limit exceeds one.

10-110

a) Underlying distributions appear to be normally distributed because the data fall along lines on the normal probability plots. The slopes appear to be similar so it is reasonable to assume that $\sigma_1^2 = \sigma_2^2$.





b)

1) The parameter of interest is the difference in mean volumes, $\mu_1 - \mu_2$

2) $H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$

3) $H_1: \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$

4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

5) Reject H_0 if $t_0 < -t_{\alpha/2,\nu}$ or $z_0 > t_{\alpha/2,\nu}$ where $t_{\alpha/2,\nu} = t_{0.025,18} = 2.101$ for $\alpha = 0.05$

6) $\bar{x}_1 = 752.7$ $\qquad \bar{x}_2 = 755.6$ $\qquad s_p = \sqrt{\dfrac{9(1.252)^2 + 9(0.843)^2}{18}} = 1.07$

$s_1 = 1.252$ $\qquad s_2 = 0.843$

$n_1 = 10$ $\quad n_2 = 10$

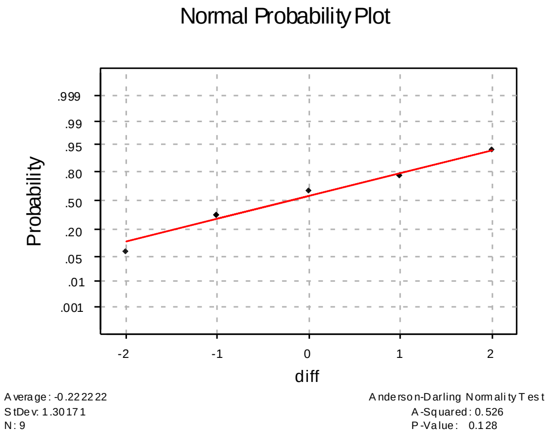
$$t_0 = \frac{(752.7 - 755.6)}{1.07\sqrt{\dfrac{1}{10} + \dfrac{1}{10}}} = -6.06$$

7) Conclusion: Because –6.06 < –2.101, reject H_0 and conclude that there is a difference between the two mean fill volumes at a 5% significance level.

c) With an estimate of $\sigma = 1.07$, $d = \Delta/2\sigma = 2/[2(1.07)] = 0.93$, and from Appendix Chart VII*e* the power is just under 80%. Because the power is relatively low, an increase in the sample size would improve the power of the test.

10-111

a) The assumption of normality appears to be reasonable. The data lie along a line in the normal probability plot.



b)
1) The parameter of interest is the mean difference in tip hardness, μ_d
2) $H_0 : \mu_d = 0$
3) $H_1 : \mu_d \neq 0$
4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

5) Since no significance level is given, we calculate the *P*-value. Reject H_0 if the *P*-value is sufficiently small.
6) $\bar{d} = -0.222$
$s_d = 1.30$
$n = 9$

$$t_0 = \frac{-0.222}{1.30 / \sqrt{9}} = -0.512$$

P-value = 2P(T < -0.512) = 2P(T > 0.512) and 2(0.25) < *P*-value < 2(0.40). Thus, 0.50 < *P*-value < 0.80

7) Conclusion: Because the P-value is greater than common levels of significance, fail to reject H_0 and conclude there is no difference in mean tip hardness.

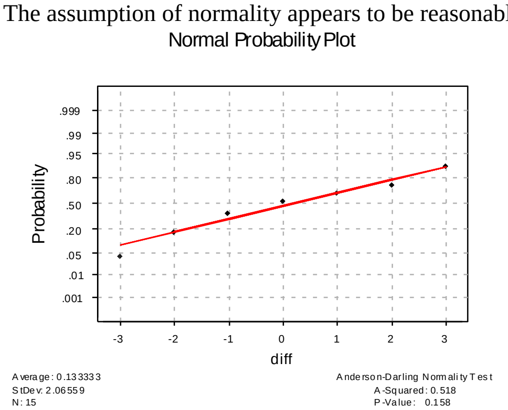
c) $\beta = 0.10$
$\mu_d = 1$

$$d = \frac{1}{\sigma_d} = \frac{1}{1.3} = 0.769$$

From Chart VII*f* with $\alpha = 0.01$, $n = 30$

10-112

a) The assumption of normality appears to be reasonable. The data lie along a line in the normal probability plot.



b)
1) The parameter of interest is the mean difference in depth using the two gauges, μ_d
2) $H_0 : \mu_d = 0$
3) $H_1 : \mu_d \neq 0$
4) The test statistic is

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

5) Because no significance level is given, we calculate the *P*-value. Reject H_0 if the *P*-value is sufficiently small.
6) $\bar{d} = 0.133$
$s_d = 2.065$
$n = 15$

$$t_0 = \frac{0.133}{2.065 / \sqrt{15}} = 0.25$$

P-value = 2P(T > 0.25), 2(0.40) < *P*-value, 0.80 < *P*-value

7) Conclusion: Because the P-value is greater than common levels of significance, fail to reject H_0. There is not sufficient evidence to conclude that the mean depth measurements for the two gauges differ at common levels of significance.

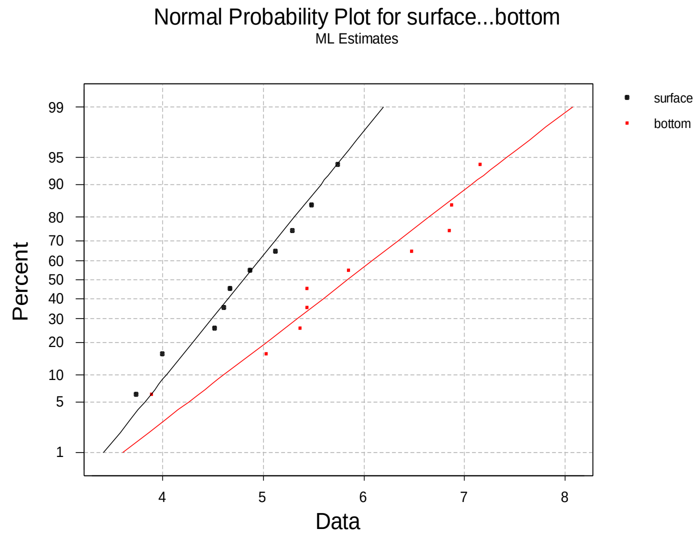
c) Power = 0.8. Because Power = 1 – β , β = 0.20
$\mu_d = 1.65$

$$d = \frac{1.65}{\sigma_d} = \frac{1.65}{(2.065)} = 0.799$$

From Chart VII*f* with α = 0.01 and β = 0.20, n = 30.

10-113

a) Because the data fall along lines, the data from both depths appear to be normally distributed, but the slopes do not appear to be equal. Therefore, it is not reasonable to assume that $\sigma_1^2 = \sigma_2^2$.



b)
1) The parameter of interest is the difference in mean HCB concentration, $\mu_1 - \mu_2$, with $\Delta_0 = 0$
2) $H_0 : \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
3) $H_1 : \mu_1 - \mu_2 \neq 0$ or $\mu_1 \neq \mu_2$
4) The test statistic is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

5) Reject the null hypothesis if $t_0 < -t_{0.025,15}$ or $t_0 > t_{0.025,15}$ where $t_{0.025,15} = 2.131$ for $\alpha = 0.05$. Also

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = 15.06$$

$$\nu \cong 15$$

6) $\bar{x}_1 = 4.804$ $\bar{x}_2 = 5.839$ $s_1 = 0.631$ $s_2 = 1.014$
$n_1 = 10$ $n_2 = 10$

$$t_0 = \frac{(4.804 - 5.839)}{\sqrt{\frac{(0.631)^2}{10} + \frac{(1.014)^2}{10}}} = -2.74$$

7) Conclusion: Because –2.74 < –2.131, reject the null hypothesis. Conclude that the mean HCB concentration is different at the two depths at a 0.05 level of significance.

c) Assume the variances are equal. Then $\Delta = 2$, $\alpha = 0.05$, $n = n_1 = n_2 = 10$, $n^* = 2n - 1 = 19$, $s_p = 0.84$

and $d = \frac{2}{2(0.84)} = 1.2$.

From Chart VII*e*, we find $\beta \approx 0.05$, and then calculate the power $= 1 - \beta = 0.95$

d) Assume the variances are equal. Then $\Delta = 1$, $\alpha = 0.05$, $n = n_1 = n_2$, $n^* = 2n - 1$, $\beta = 0.1$, $s_p = 0.84 \approx 1$

and $d = \frac{1}{2(0.84)} = 0.6$.

From Chart VII*e*, we find $n^* = 50$ and $n = \frac{50 + 1}{2} = 25.5$, so $n = 26$.

10-114 1) The parameters of interest are the foam thickness variances, σ_1^2, σ_2^2.

2) $H_0 : \sigma_1^2 = \sigma_2^2$

3) $H_1 : \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.975,8,8} = 1/4.43 = 0.226$ or $f_0 > f_{0.025,8,8} = 4.43$ for $\alpha = 0.05$

6) $n_1 = 9$ $n_2 = 9$
$s_1 = 0.049$ $s_2 = 0.094$

$$f_0 = \frac{(0.049)^2}{(0.094)^2} = 0.272$$

7) Conclusion: Because 0.226 < 0.272 < 4.43, fail to reject the null hypothesis. There is not sufficient evidence that the foam thickness variances differ at the 0.05 level of significance.

10-115 1) The parameters of interest are the grinding force variances, σ_1^2, σ_2^2.

2) $H_0 : \sigma_1^2 = \sigma_2^2$

3) $H_1 : \sigma_1^2 \neq \sigma_2^2$

4) The test statistic is

$$f_0 = \frac{s_1^2}{s_2^2}$$

5) Reject the null hypothesis if $f_0 < f_{0.975,11,11} = 1/3.474 = 0.288$ or $f_0 > f_{0.025,11,11} = 3.474$ for $\alpha = 0.05$

6) $n_1 = 12$ $\quad n_2 = 12$
$s_1 = 26.221$ $\quad s_2 = 46.596$

$$f_0 = \frac{(26.221)^2}{(46.596)^2} = 0.317$$

7) Conclusion: Because 0.288 < 0.317 < 3.474, fail to reject the null hypothesis. There is not sufficient evidence that the grinding force variances differ at the 0.05 level of significance.

10-116 95% traditional confidence interval on the difference:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(0.825-0.792)-1.96\sqrt{\frac{0.825(1-0.825)}{200}+\frac{0.792(1-0.792)}{250}} \le p_1 - p_2 \le (0.825-0.792)+1.96\sqrt{\frac{0.825(1-0.825)}{200}+\frac{0.792(1-0.792)}{250}}$$

$$-0.04 \le p_1 - p_2 \le 0.106$$

$\tilde{n}_1 = 200+2=202 \quad \tilde{n}_2 = 250+2=252$

$x_1 + 1 = 165+1=166 \qquad x_2 + 1 = 198+1=199$

$$\tilde{p}_1 = \frac{x_1+1}{n_1+1} = \frac{166}{201} = 0.83 \qquad \tilde{p}_2 = \frac{x_2+1}{n_2+1} = \frac{199}{251} = 0.793$$

95% alternate confidence interval on the difference:

$$(\tilde{p}_1 - \tilde{p}_2) - z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}} \le p_1 - p_2 \le (\tilde{p}_1 - \tilde{p}_2) + z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}}$$

$-0.035 \le p_1 - p_2 \le 0.11$.

The second interval is slightly smaller than the first interval, yet they both contain zero. Therefore, the conclusion from both intervals is that no difference is detected between the proportions at significance level 0.05.

10-117

$\tilde{n}_1 = 1500+2=1502 \qquad \tilde{n}_2 = 1200+2=1202$

$x_1 + 1 = 387+1=388 \qquad x_2 + 1 = 310+1=311$

$$\tilde{p}_1 = \frac{x_1+1}{n_1+1} = \frac{388}{1501} = 0.2585 \qquad \tilde{p}_2 = \frac{x_2+1}{n_2+1} = \frac{311}{1201} = 0.259$$

95% alternate confidence interval on the difference:

$$(\tilde{p}_1 - \tilde{p}_2) - z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}} \le p_1 - p_2 \le (\tilde{p}_1 - \tilde{p}_2) + z_{\alpha/2}\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{n}_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{\tilde{n}_2}}$$

$-0.0337 \le p_1 - p_2 \le 0.0327$

The length of the interval is close to the previous one $-0.0335 \le p_1 - p_2 \le 0.0329$ (almost the same). The conclusion from both intervals is that no difference is detected between the proportions at significance level 0.05.

Mind-Expanding Exercises

10-118

The estimate of μ is given by $\hat{\mu} = \frac{1}{2}\left(\overline{X}_1 + \overline{X}_2\right) - \overline{X}_3$. From the independence, the variance of $\hat{\mu}$ can be shown to be $V(\hat{\mu}) = \frac{1}{4}\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) + \frac{\sigma_3^2}{n_3}$.

Use s_1, s_2, and s_3 as estimates for σ_1, σ_2, and σ_3, respectively. One may also used a pooled estimate of variability.

a) An approximate 100(1 – α)% confidence interval on μ is then:

$$\hat{\mu} - Z_{\alpha/2}\sqrt{\frac{1}{4}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) + \frac{s_3^2}{n_3}} \le \mu \le \hat{\mu} + Z_{\alpha/2}\sqrt{\frac{1}{4}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) + \frac{s_3^2}{n_3}}$$

$$\left(\frac{1}{2}(4.6+5.2) - 6.1\right) - 1.96\sqrt{\frac{1}{4}\left(\frac{0.7^2}{100} + \frac{0.6^2}{120}\right) + \frac{0.8^2}{130}} \le \mu \le \left(\frac{1}{2}(4.6+5.2) - 6.1\right) + 1.96\sqrt{\frac{1}{4}\left(\frac{0.7^2}{100} + \frac{0.6^2}{120}\right) + \frac{0.8^2}{130}}$$

$$-1.2 - 0.163 \le \mu \le -1.2 + 0.163$$

$$-1.363 \le \mu \le -1.037$$

b) An approximate one-sided 95% confidence interval for $\hat{\mu}$ is

$$\mu \le \left(\frac{1}{2}(4.6+5.2) - 6.1\right) + 1.64\sqrt{\frac{1}{4}\left(\frac{0.7^2}{100} + \frac{0.6^2}{120}\right) + \frac{0.8^2}{130}}$$

$$\mu \le -1.2 + 0.136$$

$$\mu \le -1.064$$

Because the interval is negative and does not contain zero, we can conclude that that pesticide three is more effective.

10-119 The $V(\overline{X}_1 - \overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ and suppose this is to equal a constant k. Then, we are to minimize $C_1 n_1 + C_2 n_2$ subject to $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = k$. Using a Lagrange multiplier, we minimize by setting the partial derivatives of $f(n_1, n_2, \lambda) = C_1 n_1 + C_2 n_2 + \lambda\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} - k\right)$ with respect to n_1, n_2 and λ equal to zero.

These equations are

$$\frac{\partial}{\partial n_1} f(n_1, n_2, \lambda) = C_1 - \frac{\lambda\sigma_1^2}{n_1^2} = 0 \qquad (1)$$

$$\frac{\partial}{\partial n_2} f(n_1, n_2, \lambda) = C_2 - \frac{\lambda\sigma_2^2}{n_2^2} = 0 \qquad (2)$$

$$\frac{\partial}{\partial \lambda} f(n_1, n_2, \lambda) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = k \qquad (3)$$

Upon adding equations (1) and (2), we obtain $C_1 + C_2 - \lambda\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) = 0$

Substituting from equation (3) enables us to solve for λ to obtain $\frac{C_1 + C_2}{k} = \lambda$

Then, equations (1) and (2) are solved for n_1 and n_2 to obtain

$$n_1 = \frac{\sigma_1^2(C_1 + C_2)}{kC_1} \qquad n_2 = \frac{\sigma_2^2(C_1 + C_2)}{kC_2}$$

It can be verified that this is a minimum. With these choices for n_1 and n_2

$$V(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

10-120 Maximizing the probability of rejecting H_0 is equivalent to minimizing

$$P\left(-z_{\alpha/2} < \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < z_{\alpha/2} \mid \mu_1 - \mu_2 = \delta\right) = P\left(-z_{\alpha/2} - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < Z < z_{\alpha/2} - \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right)$$

where z is a standard normal random variable. This probability is minimized by maximizing $\frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$.

Therefore, we are to minimize $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ subject to $n_1 + n_2 = N$.

From the constraint, $n_2 = N - n_1$, we are to minimize $f(n_1) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{N - n_1}$.

Take the derivative of $f(n_1)$ with respect to n_1 and set it equal to zero results in the equation $\frac{-\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{(N-n_1)^2} = 0$.

Solve for n_1 to obtain $n_1 = \frac{\sigma_1 N}{\sigma_1 + \sigma_2}$ and $n_2 = \frac{\sigma_2 N}{\sigma_1 + \sigma_2}$

Also, it can be verified that the solution minimizes $f(n_1)$.

10-121

a) $\alpha = P(Z > z_\varepsilon \text{ or } Z < -z_{\alpha-\varepsilon})$ where Z has a standard normal distribution.

Then, $\alpha = P(Z > z_\varepsilon) + P(Z < -z_{\alpha-\varepsilon}) = \varepsilon + \alpha - \varepsilon = \alpha$

b) $\beta = P(-z_{\alpha-\varepsilon} < Z_0 < z_\varepsilon | \mu_1 = \mu_0 + \delta)$

$$\begin{aligned}\beta &= P(-z_{\alpha-\varepsilon} < \tfrac{\bar{x}-\mu_0}{\sqrt{\sigma^2/n}} < z_\varepsilon \mid \mu_1 = \mu_0 + \delta) \\ &= P(-z_{\alpha-\varepsilon} - \tfrac{\delta}{\sqrt{\sigma^2/n}} < Z < z_\varepsilon - \tfrac{\delta}{\sqrt{\sigma^2/n}}) \\ &= \Phi(z_\varepsilon - \tfrac{\delta}{\sqrt{\sigma^2/n}}) - \Phi(-z_{\alpha-\varepsilon} - \tfrac{\delta}{\sqrt{\sigma^2/n}})\end{aligned}$$

10-122 The requested result can be obtained from data in which the pairs are very different. Example:

pair	1	2	3	4	5
sample 1	100	10	50	20	70
sample 2	110	20	59	31	80

$\bar{x}_1 = 50$ $\bar{x}_2 = 60$

$s_1 = 36.74$ $s_2 = 36.54$ $s_{pooled} = 36.64$

Two-sample t-test : $t_0 = -0.43$ *P*-value = 0.68

$\bar{x}_d = -10$ $s_d = 0.707$

Paired t-test: $t_0 = -3162$ *P*-value ≈ 0

10-123 a) $\theta = \frac{p_1}{p_2}$ and $\hat{\theta} = \frac{\hat{p}_1}{\hat{p}_2}$ and $\ln(\hat{\theta}) \sim N[\ln(\theta), \sqrt{(n_1 - x_1)/n_1 x_1 + (n_2 - x_2)/n_2 x_2}\,]$

The $(1-\alpha)$ confidence interval for $\ln(\theta)$ can use the relationship $Z = \dfrac{\ln(\hat{\theta}) - \ln(\theta)}{\left(\left(\dfrac{n_1 - x_1}{n_1 x_1}\right) + \left(\dfrac{n_2 - x_2}{n_2 x_2}\right)\right)^{1/4}}$

$$\ln(\hat{\theta}) - Z_{\alpha/2}\left(\left(\frac{n_1 - x_1}{n_1 x_1}\right) + \left(\frac{n_2 - x_2}{n_2 x_2}\right)\right)^{1/4} \le \ln(\theta) \le \ln(\hat{\theta}) + Z_{\alpha/2}\left(\left(\frac{n_1 - x_1}{n_1 x_1}\right) + \left(\frac{n_2 - x_2}{n_2 x_2}\right)\right)^{1/4}$$

b) The $(1-\alpha)$ confidence interval for θ can use the CI developed in part (a) where θ = e^(ln(θ))

$$\hat{\theta} e^{-Z_{\alpha/2}\left(\left(\frac{n_1 - x_1}{n_1 x_1}\right) + \left(\frac{n_2 - x_2}{n_2 x_2}\right)\right)^{1/4}} \le \theta \le \hat{\theta} e^{Z_{\alpha/2}\left(\left(\frac{n_1 - x_1}{n_1 x_1}\right) + \left(\frac{n_2 - x_2}{n_2 x_2}\right)\right)^{1/4}}$$

c)

$$\hat{\theta} e^{-Z_{\alpha/2}\left(\left(\frac{n_1 - x_1}{n_1 x_1}\right) + \left(\frac{n_2 - x_2}{n_2 x_2}\right)\right)^{.25}} \le \theta \le \hat{\theta} e^{Z_{\alpha/2}\left(\left(\frac{n_1 - x_1}{n_1 x_1}\right) + \left(\frac{n_2 - x_2}{n_2 x_2}\right)\right)^{.25}}$$

$$1.42 e^{-1.96\left(\left(\frac{100-27}{2700}\right) + \left(\frac{100-19}{1900}\right)\right)^{1/4}} \le \theta \le 1.42 e^{1.96\left(\left(\frac{100-27}{2700}\right) + \left(\frac{100-19}{1900}\right)\right)^{1/4}}$$

$$0.519 \le \theta \le 3.887$$

Because the confidence interval contains the value one, we conclude that there is no significant difference in the proportions at the 95% level of significance.

10-124 $H_0 : \sigma_1^2 = \sigma_2^2$

$$H_1 : \sigma_1^2 \ne \sigma_2^2$$

$$\beta = P\left(f^{2}_{1-\alpha/2, n_1-1, n_2-1} < \frac{S_1^2}{S_2^2} < f^{2}_{\alpha/2, n_1-1, n_2-1} \mid \frac{\sigma_1^2}{\sigma_2^2} = \delta \ne 1 \right)$$

$$= P\left(\frac{\sigma_2^2}{\sigma_1^2} f_{1-\alpha/2, n_1-1, n_2-1} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < \frac{\sigma_2^2}{\sigma_1^2} f_{\alpha/2, n_1-1, n_2-1} \mid \frac{\sigma_1^2}{\sigma_2^2} = \delta \right)$$

where $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ has an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

CHAPTER 11

Section 11-2

11-1 a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 162674.2 - \tfrac{6322.28^2}{250} = 2789.282$$

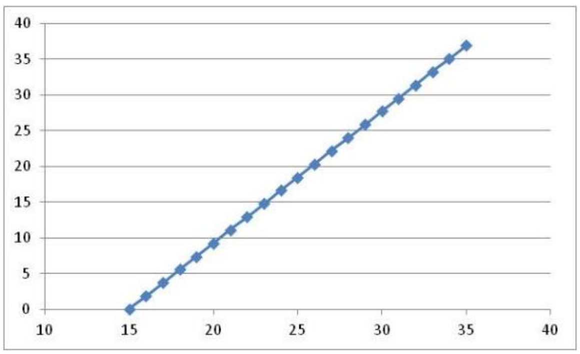
$$S_{xy} = 125471.1 - \tfrac{(6322.28)(4757.9)}{250} = 5147.996$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{5147.996}{2789.282} = 1.846$$

$$\hat{\beta}_0 = 19.032 - (1.846)(25.289) = -27.643$$

$$\hat{y} = -27.643 + 1.846x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{17128.82 - 1.846(5147.996)}{248} = 30.756$$



b) $\hat{y} = -27.643 + 1.846(30) = 27.726$

c) $\hat{y} = -27.643 + 1.846(25) = 18.498$
residual = 25 – 18.498 = 6.50

d) Because the actual BMI is greater than the prediction (residual is positive), this is an underestimate of the BMI

11-2 a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 543503 - \tfrac{11211^2}{250} = 40756.92$$

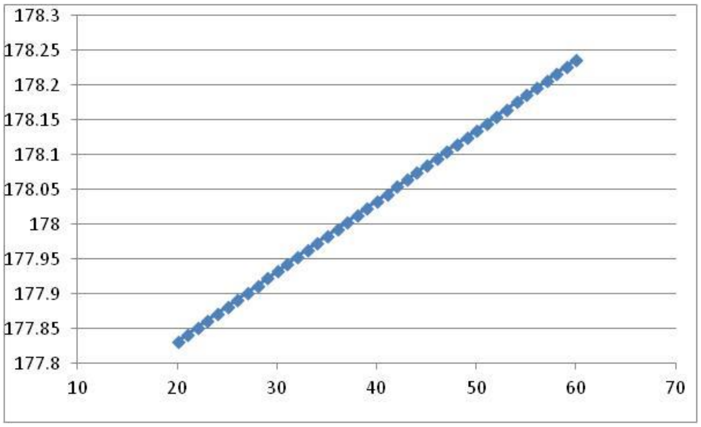
$$S_{xy} = 1996904.15 - \tfrac{(611211)(44520.8)}{250} = 413.395$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{413.395}{40756.92} = 0.01014$$

$$\hat{\beta}_0 = 178.083 - (0.01014)(44.844) = 177.628$$

$$\hat{y} = 177.628 + 0.01014x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{181998.489 - 0.01014(413.395)}{248} = 733.848$$



b) $\hat{y} = 177.628 + 0.01014(25) = 177.882$

c) $\hat{y} = 177.628 + 0.01014(25) = 177.882$
residual = 170 – 177.882 = -7.88

d) Because the actual BMI is less than the prediction (residual is negative), this is an overestimate of the BMI

11-3 a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 157.42 - \tfrac{43^2}{14} = 25.348571$$

$$S_{xy} = 1697.80 - \tfrac{43(572)}{14} = -59.057143$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-59.057143}{25.348571} = -2.330$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \tfrac{572}{14} - (-2.3298017)(\tfrac{43}{14}) = 48.013$$

$$SS_R = \hat{\beta}_1 S_{xy} = -2.3298017(-59.057143)$$
$$= 137.59$$

$$SS_E = S_{yy} - SS_R$$
$$= 159.71429 - 137.59143$$
$$= 22.123$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{22.123}{12} = 1.8436$$

b) $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, $\hat{y} = 48.012962 - 2.3298017(4.3) = 37.99$

c) $\hat{y} = 48.012962 - 2.3298017(3.7) = 39.39$

d) $e = y - \hat{y} = 46.1 - 39.39 = 6.71$

11-4 a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 143215.8 - \tfrac{1478^2}{20} = 33991.6$$

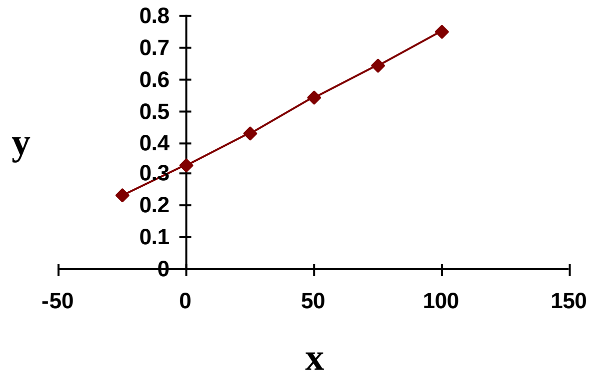
$$S_{xy} = 1083.67 - \tfrac{(1478)(12.75)}{20} = 141.445$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{141.445}{33991.6} = 0.00416$$

$$\hat{\beta}_0 = \tfrac{12.75}{20} - (0.0041617512)(\tfrac{1478}{20}) = 0.32999$$

$$\hat{y} = 0.32999 + 0.00416x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{0.143275}{18} = 0.00796$$



b) $\hat{y} = 0.32999 + 0.00416(85) = 0.6836$

c) $\hat{y} = 0.32999 + 0.00416(90) = 0.7044$

d) $\hat{\beta}_1 = 0.00416$

11-5 a)

Regression Analysis: Rating Pts versus Yds per Att

```
The regression equation is
Rating Pts = 14.2 + 10.1 Yds per Att

Predictor       Coef  SE Coef     T      P
Constant      14.195    9.059  1.57  0.128
Yds per Att   10.092    1.288  7.84  0.000

S = 5.21874   R-Sq = 67.2%   R-Sq(adj) = 66.1%

Analysis of Variance
Source          DF      SS      MS      F      P
Regression       1  1672.5  1672.5  61.41  0.000
Residual Error  30   817.1    27.2
Total           31  2489.5
```

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$S_{xx} = 1583.442 - \frac{(223.93)^2}{32} = 16.422$$

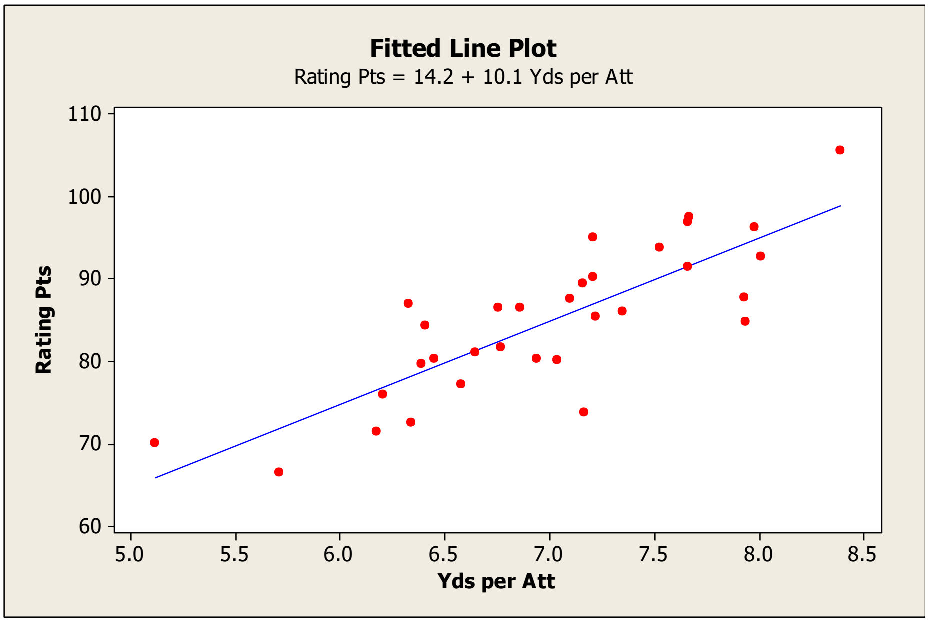
$$S_{xy} = 19158.49 - \frac{(223.93)(2714.1)}{32} = 165.7271$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{165.7271}{16.422} = 10.092$$

$$\hat{\beta}_0 = \frac{2714.1}{32} - (10.092)\left(\frac{223.93}{32}\right) = 14.195$$

$$\hat{y} = 14.2 + 10.1x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{817.1}{30} = 27.24$$



b) $\hat{y} = 14.2 + 10.1(7.5) = 89.95$

c) $-\hat{\beta}_1 = -10.1$

d) $\dfrac{1}{10.1} \times 10 = 0.99$

e) $\hat{y} = 14.2 + 10.1(7.21) = 87.02$

There are two residuals

$$e = y - \hat{y}$$

$$e_1 = 90.2 - 87.02 = 3.18$$

$$e_2 = 95 - 87.02 = 7.98$$

11-6 a)

```
Regression Analysis - Linear model: Y = a+bX
Dependent variable: SalePrice                    Independent variable: Taxes
--------------------------------------------------------------------------------
                               Standard              T              Prob.
Parameter       Estimate         Error             Value            Level
Intercept        13.3202        2.57172           5.17948          .00003
Slope            3.32437       0.390276             8.518          .00000
--------------------------------------------------------------------------------
                           Analysis of Variance
Source             Sum of Squares     Df  Mean Square    F-Ratio  Prob. Level
Model                   636.15569      1    636.15569    72.5563       .00000
Residual                192.89056     22      8.76775
--------------------------------------------------------------------------------
Total (Corr.)           829.04625     23
Correlation Coefficient = 0.875976          R-squared =  76.73 percent
Stnd. Error of Est. = 2.96104
```

$$\hat{\sigma}^2 = 8.76775$$

$$\hat{y} = 13.3202 + 3.32437x$$

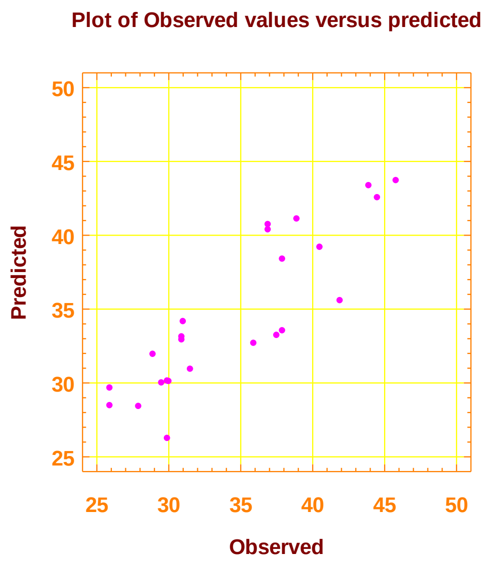
b) $\hat{y} = 13.3202 + 3.32437(7.5) = 38.253$

c) $\hat{y} = 13.3202 + 3.32437(5.8980) = 32.9273$

$$\hat{y} = 32.9273$$

$$e = y - \hat{y} = 30.9 - 32.9273 = -2.0273$$

d) All the points would lie along a 45 degree line. That is, the regression model would estimate the values exactly. At this point, the graph of observed vs. predicted indicates that the simple linear regression model provides a reasonable fit to the data.



11-7 a)

```
Regression Analysis - Linear model: Y = a+bX
Dependent variable: Usage              Independent variable: Temperature
--------------------------------------------------------------------------------
                               Standard            T            Prob.
Parameter       Estimate         Error           Value          Level
Intercept        -6.3355        1.66765        -3.79906        .00349
Slope            9.20836      0.0337744        272.643         .00000
--------------------------------------------------------------------------------
                           Analysis of Variance
Source             Sum of Squares     Df  Mean Square    F-Ratio  Prob. Level
Model                   280583.12      1    280583.12    74334.4       .00000
Residual                37.746089     10     3.774609
--------------------------------------------------------------------------------
Total (Corr.)           280620.87     11
Correlation Coefficient = 0.999933          R-squared =  99.99 percent
Stnd. Error of Est. = 1.94284
```

$\hat{\sigma}^2 = 3.7746$

$\hat{y} = -6.3355 + 9.20836x$

b) $\hat{y} = -6.3355 + 9.20836(55) = 500.124$

c) If monthly temperature increases by 1°F, $\hat{y}$ increases by 9.208

d) $\hat{y} = -6.3355 + 9.20836(47) = 426.458$

$\hat{y} = 426.458$

$e = y - \hat{y} = 424.84 - 426.458 = -1.618$

11-8 a)

```
The regression equation is
MPG = 39.2 - 0.0402 Engine Displacement

Predictor                Coef   SE Coef      T      P
Constant               39.156     2.006  19.52  0.000
Engine Displacement  -0.040216  0.007671  -5.24  0.000

S = 3.74332   R-Sq = 59.1%   R-Sq(adj) = 57.0%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  385.18  385.18  27.49  0.000
Residual Error  19  266.24   14.01
Total           20  651.41
```

$\hat{\sigma}^2 = 14.01$

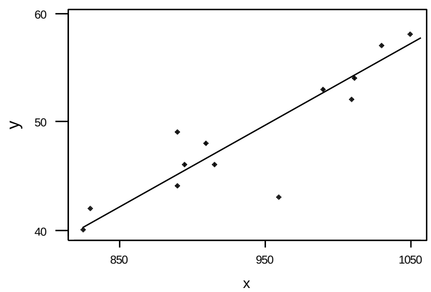
$\hat{y} = 39.2 - 0.0402x$

b) $\hat{y} = 39.2 - 0.0402(150) = 33.17$

c) $\hat{y} = 34.2956$

$e = y - \hat{y} = 41.3 - 34.2956 = 7.0044$

11-9 a)




```
Predictor       Coef       StDev          T        P
Constant     -16.509       9.843      -1.68    0.122
x            0.06936     0.01045       6.64    0.000
S = 2.706      R-Sq = 80.0%     R-Sq(adj) = 78.2%
Analysis of Variance
Source       DF         SS          MS         F        P
Regression    1     322.50      322.50     44.03    0.000
Error        11      80.57        7.32
Total        12     403.08
```

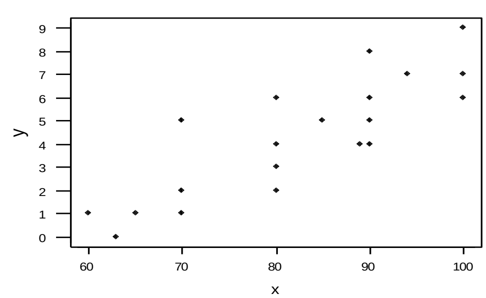
$$\hat{\sigma}^2 = 7.3212$$

$$\hat{y} = -16.5093 + 0.0693554x$$

b) $\hat{y} = 46.6041$ $\quad e = y - \hat{y} = 1.39592$

c) $\hat{y} = -16.5093 + 0.0693554(950) = 49.38$

11-10 a)



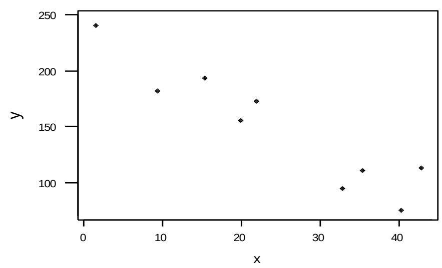
Yes, a linear regression seems appropriate, but one or two points might be outliers.

```
Predictor        Coef     SE Coef          T        P
Constant      -10.132       1.995      -5.08    0.000
x             0.17429     0.02383       7.31    0.000
S = 1.318      R-Sq = 74.8%     R-Sq(adj) = 73.4%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         1      92.934      92.934     53.50    0.000
Residual Error    18      31.266       1.737
Total             19     124.200
```

b) $\hat{\sigma}^2 = 1.737$ and $\hat{y} = -10.132 + 0.17429x$

c) $\hat{y} = 4.68265$ at $x = 85$

11-11 a)



Yes, a linear regression model appears to be plausible.

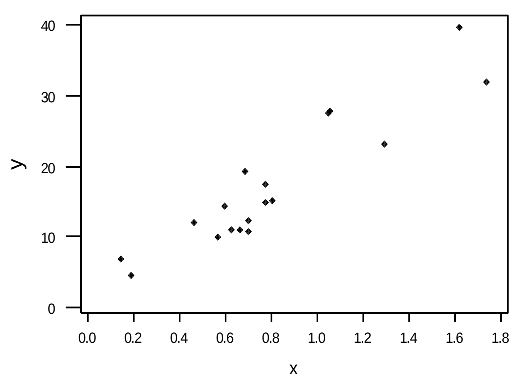
```
Predictor        Coef       StDev          T        P
Constant       234.07       13.75      17.03    0.000
x             -3.5086      0.4911      -7.14    0.000
S = 19.96       R-Sq = 87.9%     R-Sq(adj) = 86.2%
Analysis of Variance
Source        DF          SS          MS         F        P
Regression     1       20329       20329     51.04    0.000
Error          7        2788         398
Total          8       23117
```

b) $\hat{\sigma}^2 = 398.25$ and $\hat{y} = 234.071 - 3.50856x$

c) $\hat{y} = 234.071 - 3.50856(30) = 128.814$

d) $\hat{y} = 156.883 \quad e = 15.1175$

11-12 a)



Yes, a simple linear regression model seems appropriate for these data.

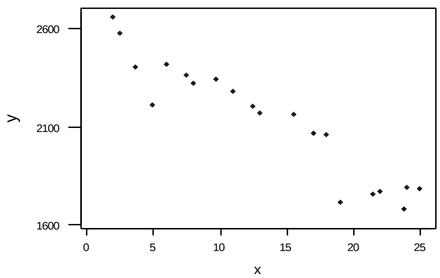
```
Predictor        Coef       StDev          T        P
Constant        0.470       1.936       0.24    0.811
x              20.567       2.142       9.60    0.000
S = 3.716       R-Sq = 85.2%     R-Sq(adj) = 84.3%

Analysis of Variance

Source        DF          SS          MS         F        P
Regression     1      1273.5      1273.5     92.22    0.000
Error         16       220.9        13.8
Total         17      1494.5
```

b) $\hat{\sigma}^2 = 13.81$
$\hat{y} = 0.470467 + 20.5673x$

c) $\hat{y} = 0.470467 + 20.5673(1) = 21.038$

d) $\hat{y} = 10.1371 \quad e = 1.6629$

11-13 a)



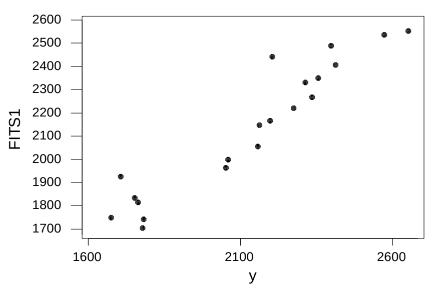
Yes, a simple linear regression model seems plausible for this situation.

```
Predictor        Coef       StDev           T         P
Constant      2625.39       45.35       57.90     0.000
x             -36.962       2.967      -12.46     0.000

S = 99.05       R-Sq = 89.6%     R-Sq(adj) = 89.0%

Analysis of Variance
Source        DF          SS          MS          F         P
Regression     1     1522819     1522819     155.21     0.000
Error         18      176602        9811
Total         19     1699421
```

b) $\hat{\sigma}^2 = 9811.2$
$\hat{y} = 2625.39 - 36.962x$

c) $\hat{y} = 2625.39 - 36.962(20) = 1886.15$

d) If there were no error, the values would all lie along the 45 $^\circ$ line. The plot indicates age is reasonable regressor variable.



11-14 a)

```
The regression equation is
Porosity = 55.6 - 0.0342 Temperature

Predictor         Coef  SE Coef      T      P
Constant         55.63    32.11   1.73  0.144
Temperature   -0.03416  0.02569  -1.33  0.241

S = 8.79376   R-Sq = 26.1%   R-Sq(adj) = 11.3%

Analysis of Variance

Source          DF      SS      MS     F      P
Regression       1  136.68  136.68  1.77  0.241
Residual Error   5  386.65   77.33
Total            6  523.33
```

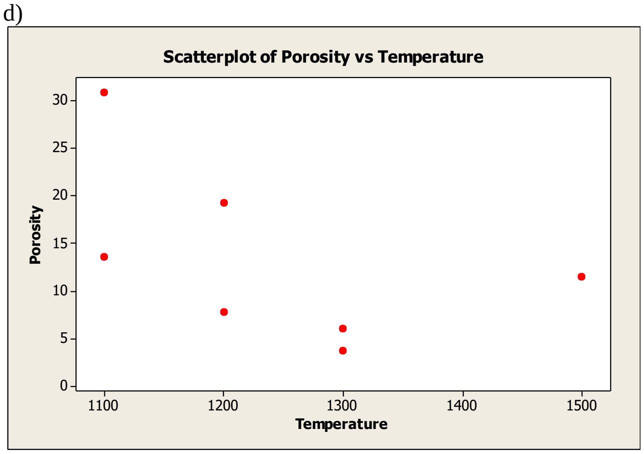
$$\hat{y} = 55.63 - 0.03416x$$

$$\hat{\sigma}^2 = 77.33$$

b) $\hat{y} = 55.63 - 0.03416(1400) = 7.806$

c) $\hat{y} = 4.39 \quad e = 7.012$

d)



The simple linear regression model doesn't seem appropriate because the scatter plot doesn't indicate a linear relationship.

11-15 a)

```
The regression equation is
BOD = 0.658 + 0.178 Time

Predictor      Coef  SE Coef      T      P
Constant     0.6578   0.1657   3.97  0.003
Time        0.17806  0.01400  12.72  0.000

S = 0.287281   R-Sq = 94.7%   R-Sq(adj) = 94.1%

Analysis of Variance
Source          DF      SS      MS       F      P
Regression       1  13.344  13.344  161.69  0.000
Residual Error   9   0.743   0.083
Total           10  14.087
```

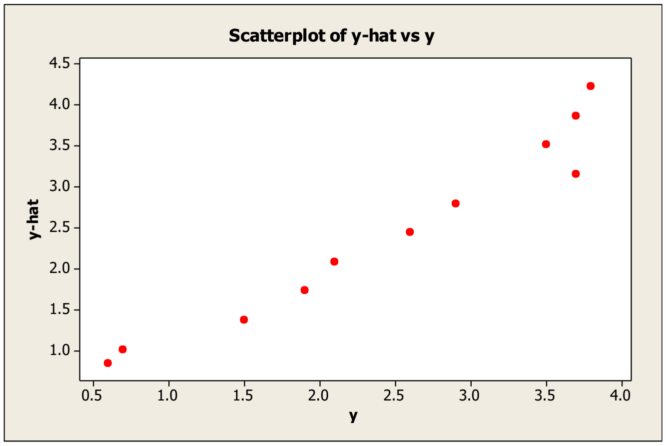
$\hat{y} = 0.658 + 0.178x$
$\hat{\sigma}^2 = 0.083$

b) $\hat{y} = 0.658 + 0.178(15) = 3.328$

c) 0.178(3) = 0.534

d) $\hat{y} = 0.658 + 0.178(6) = 1.726$
$e = y - \hat{y} = 1.9 - 1.726 = 0.174$

e)

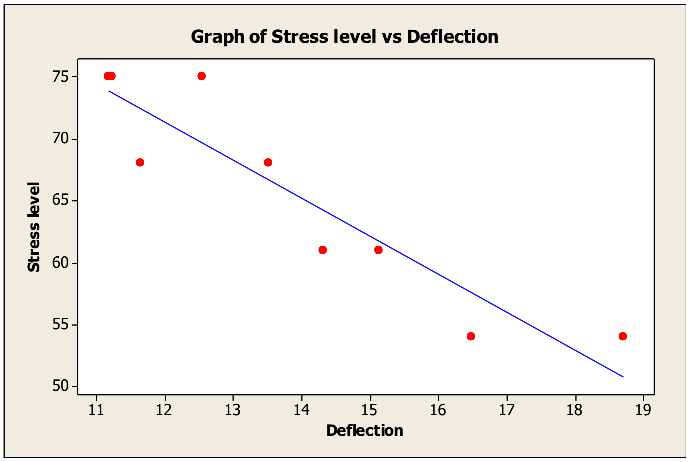


All the points would lie along the 45 degree line $y = \hat{y}$. That is, the regression model would estimate the values exactly. At this point, the graph of observed vs. predicted indicates that the simple linear regression model provides a reasonable fit to the data.

11-16 a)

```
The regression equation is
Deflection = 32.0 - 0.277 Stress level

Predictor          Coef   SE Coef      T      P
Constant         32.049     2.885  11.11  0.000
Stress level   -0.27712   0.04361  -6.35  0.000

S = 1.05743   R-Sq = 85.2%   R-Sq(adj) = 83.1%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  45.154  45.154  40.38  0.000
Residual Error   7   7.827   1.118
Total            8  52.981
```

$\hat{\sigma}^2 = 1.118$



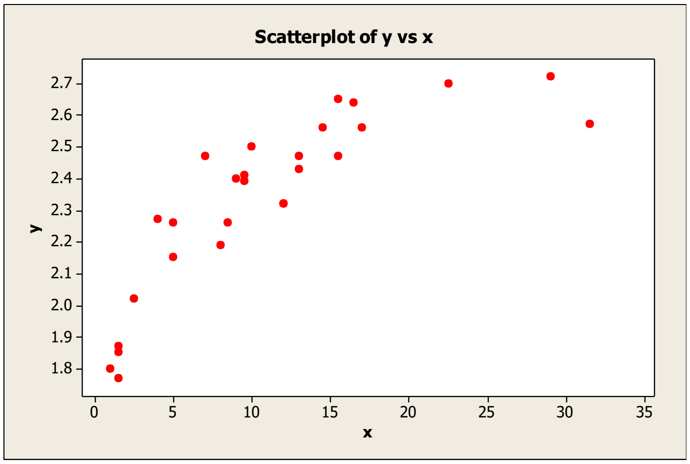
b) $\hat{y} = 32.05 - 0.277(65) = 14.045$

c) (-0.277)(5) = -1.385

d) $\dfrac{1}{0.277} = 3.61$

e) $\hat{y} = 32.05 - 0.277(68) = 13.214 \;\; e = y - \hat{y} = 11.640 - 13.214 = 1.574$

11-17



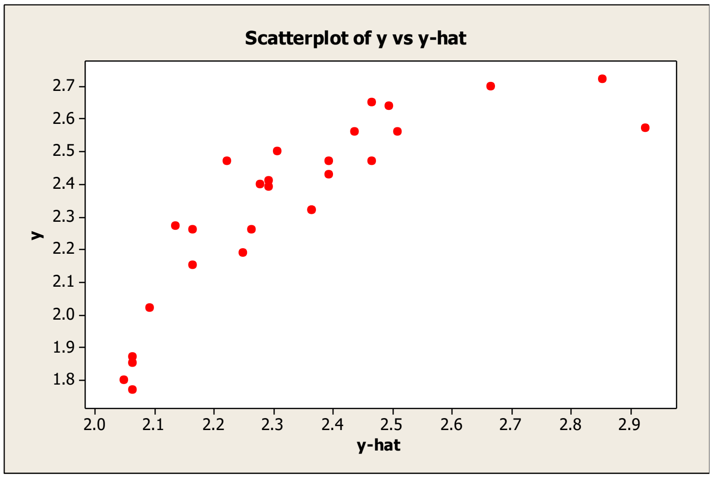
It's possible to fit this data with linear model, but it's not a good fit. Curvature is seen on the scatter plot.

a)
```
The regression equation is
y = 2.02 + 0.0287 x

Predictor       Coef    SE Coef      T      P
Constant     2.01977    0.05313  38.02  0.000
x           0.028718   0.003966   7.24  0.000

S = 0.159159   R-Sq = 67.7%   R-Sq(adj) = 66.4%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  1.3280  1.3280  52.42  0.000
Residual Error  25  0.6333  0.0253
Total           26  1.9613
```

$$\hat{y} = 2.02 + 0.0287x$$
$$\hat{\sigma}^2 = 0.0253$$

b) $\hat{y} = 2.02 + 0.0287(11) = 2.3357$

c)



If the relationship between length and age was deterministic, the points would fall on the 45 degree line $y = \hat{y}$. The plot does not indicate a linear relationship. Therefore, age is not a reasonable choice for the regressor variable in this model.

11-18 a) $\hat{y} = 0.3299892 + 0.0041612(\frac{9}{5}x + 32)$
$$\hat{y} = 0.3299892 + 0.0074902x + 0.1331584$$
$$\hat{y} = 0.4631476 + 0.0074902x$$
b) $\hat{\beta}_1 = 0.00749$

11-19 Let x = engine displacement (cm^3) and x_{old} = engine displacement (in^3)

a) The old regression equation is $y = 39.2 - 0.0402x_{old}$

Because 1 in^3 = 16.387 cm^3, the new regression equation is
$$\hat{y} = 39.2 - 0.0402(x/16.387) = 39.2 - 0.0025x$$
b) $\hat{\beta}_1 = -0.0025$

11-20 $\hat{\beta}_0 + \hat{\beta}_1\bar{x} = (\bar{y} - \hat{\beta}_1\bar{x}) + \hat{\beta}_1\bar{x} = \bar{y}$

11-21 a) The slopes of both regression models will be the same, but the intercept will be shifted.
b) $\hat{y} = 2132.41 - 36.9618x$

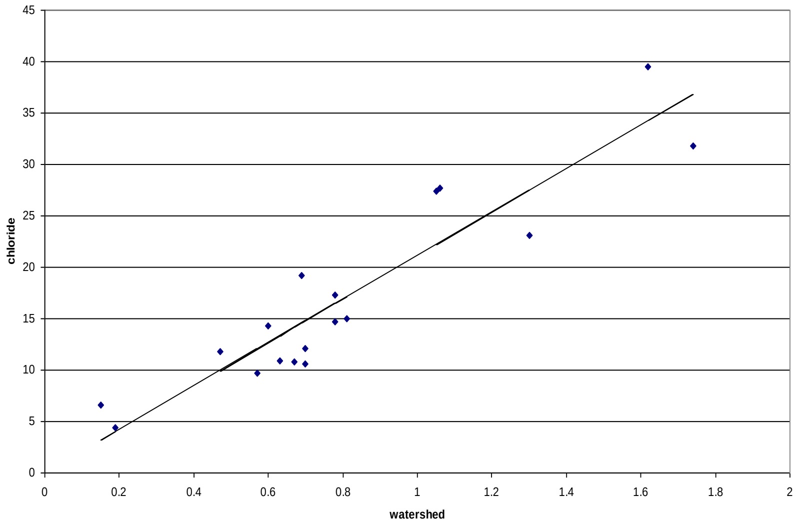
$$\hat{\beta}_0 = 2625.39 \qquad \text{vs.} \qquad \hat{\beta}_0^* = 2132.41$$
$$\hat{\beta}_1 = -36.9618 \qquad\qquad \hat{\beta}_1^* = -36.9618$$

11-22 a) The least squares estimate minimizes $\sum (y_i - \beta x_i)^2$. Upon setting the derivative equal to zero, we obtain

$$2\sum (y_i - \beta x_i)\ (-x_i) = 2[\sum - y_i x_i + \beta \sum x_i^{\ 2}] = 0$$

Therefore, $\hat{\beta} = \dfrac{\sum y_i x_i}{\sum x_i^{\ 2}}$.

b) $\hat{y} = 21.031461x$. The model seems very appropriate—an even better fit.



Section 11-4

11-23 a) $\hat{\sigma}^2 = MS_E = \dfrac{SS_E}{n-2} = \dfrac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \dfrac{17128.82 - 1.846(5147.10)}{248} = 30.756$

and $\hat{\sigma} = 5.546$

b) $se\hat{\beta}_1 = \sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\dfrac{30.756}{2789.282}} = 0.105$

c) $t = \dfrac{\hat{\beta}_1}{se\hat{\beta}_1} = \dfrac{1.846}{0.105} = 17.58$

d) P-value = P(|t_{248}| > 17.58) ≈ 0, reject H_0

11-24 a) $\hat{\sigma}^2 = MS_E = \dfrac{SS_E}{n-2} = \dfrac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \dfrac{1181998.489 - 0.01014(413.395)}{248} = 733.848$

and $\hat{\sigma} = 27.090$

b) $se\hat{\beta}_1 = \sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\dfrac{27.090}{40756.916}} = 0.134$

c) $t = \dfrac{\hat{\beta}_1}{se\hat{\beta}_1} = \dfrac{0.01014}{0.134} = 0.076$

d) P-value = P(|t_{248}| > 0.076) ≈ 1, fail to reject H_0

11-25 1 kilogram = 2.20462 pounds

a) Because the y's are divided by 2.20462 pounds, the intercept and slope estimates are divided by the same value.

$$\hat{\beta}_1 = \frac{0.01014}{2.20462} = 0.0046$$

$$\hat{\beta}_0 = \frac{177.628}{2.20462} = 80.571$$

b) Because the y's are divided by 2.20462 pounds, the error standard deviation is also divided by the same value

$$\hat{\sigma} = \frac{27.090}{2.20462} = 12.288$$

c) Because the y's are divided by 2.20462 pounds, the estimate of the standard deviation of the slope is also divided by the same value

$$se\hat{\beta}_1 = \frac{0.134}{2.20462} = 0.061$$

d) Because the estimate of the slope and the estimate of the standard error of the slope are both divided by 2.20462 pounds, the t statistic does not change

$$t = \frac{\hat{\beta}_1}{se\hat{\beta}_1} = \frac{0.01014}{0.134} = 0.076$$

e) Because the t statistic does not change, the conclusions from the test do not change
P-value = $P(|t_{248}| > 0.076) \approx 1$, fail to reject H_0

11-26 Results will depend on the simulated errors. Example data are shown.

x	e	y
10	0.159207	260.159207
12	-0.8548	309.1451978
14	3.040714	363.0407136
16	-0.05424	409.9457555
18	-2.57662	457.4233788
20	3.050463	513.0504634
22	-0.02042	559.979575
24	2.404245	612.4042447

a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 2480 - \tfrac{136^2}{8} = 168$$

$$S_{xy} = 63464.919 - \tfrac{(136)(3485.149)}{8} = 4217.394$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{4217.394}{168} = 25.104$$

$$\hat{\beta}_0 = 435.644 - (25.104)(17) = 8.883$$

$$\hat{y} = 8.883 + 25.104x$$

b) $\hat{\sigma}^2 = MS_E = \dfrac{SS_E}{n-2} = \dfrac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \dfrac{105898.118 - 25.104(4217.394)}{6} = 4.436$

$$\hat{\sigma} = 2.106$$

c) $se\hat{\beta}_1 = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{4.436}{168}} = 0.162$

$$se\hat{\beta}_0 = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\overline{x^2}}{S_{xx}}\right)} = \sqrt{4.436\left(\frac{1}{8} + \frac{17^2}{168}\right)} = 2.861$$

d) Results will depend on the simulated data. Example data are shown.

x	e	y
10	-3.81589	256.1841058
12	1.792464	311.792464
14	3.022851	363.0228512
16	-2.31405	407.6859453
18	-0.25721	459.7427926
20	0.973364	510.973364
22	-1.97236	558.0276425
24	-1.82795	608.1720475
10	-1.29204	258.7079582
12	1.181634	311.1816337
14	-0.35598	359.644017
16	2.910692	412.9106923
18	-0.7661	459.2338962
20	0.077666	510.0776663
22	0.79119	560.7911899
24	-3.91594	606.0840614

e) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 4960 - \tfrac{272^2}{16} = 336$$

$$S_{xy} = 126590.254 - \tfrac{(272)(6954.232)}{16} = 8368.305$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{8368.305}{336} = 24.906$$

$$\hat{\beta}_0 = 434.640 - (24.906)(17) = 11.243$$

$$\hat{y} = 11.243 + 24.906x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{208481.877 - 24.906(8368.305)}{14} = 4.547$$

$$\hat{\sigma} = 2.132$$

The estimate of σ is similar to the case with *n* = 8.

f) $se\hat{\beta}_1 = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{4.547}{336}} = 0.116$

$$se\hat{\beta}_0 = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\overline{x^2}}{S_{xx}}\right)} = \sqrt{4.547\left(\frac{1}{16} + \frac{17^2}{336}\right)} = 2.084$$

The standard error estimates of $\hat{\beta}_0, \hat{\beta}_1$ are both reduced by the larger sample size.

11-27 a) $T_0 = \dfrac{\hat{\beta}_0 - \beta_0}{se(\beta_0)} = \dfrac{12.857}{1.032} = 12.4583$

P-value = 2[P(T_8 > 12.4583)] and P-value < 2(0.0005) = 0.001

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{se(\beta_1)} = \frac{2.3445}{0.115} = 20.387$$

P-value = 2[P(T_8 > 20.387)] and P-value < 2(0.0005) = 0.001

$$MS_E = \frac{SS_E}{n-2} = \frac{17.55}{8} = 2.1938$$

$$F_0 = \frac{MS_R}{MS_E} = \frac{912.43}{2.1938} = 415.913$$

P-value is near zero

b) Because the P-value of the F-test ≈ 0 is less than α = 0.05, we reject the null hypothesis that $\beta_1 = 0$ at the 0.05 level of significance. This is the same result obtained from the T_1 test. If the assumptions are valid, a useful linear relationship exists.

c) $\hat{\sigma}^2 = MS_E = 2.1938$

11-28 a) $T_0 = \dfrac{\hat{\beta}_0 - \beta_0}{se(\beta_0)} = \dfrac{26.753}{2.373} = 11.2739$

P-value = 2[P(T_{14} > 11.2739)] and P-value < 2(0.0005) = 0.001

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{se(\beta_1)} = \frac{1.4756}{0.1063} = 13.8815$$

P-value = 2[P(T_{14} > 13.8815)] and P-value < 2(0.0005) = 0.001

Degrees of freedom of the residual error = 15 – 1 = 14.
Sum of squares regression = Sum of square Total – Sum of square residual error = 1500 – 94.8 = 1405.2

$$MS_{\mathrm{Re}gression} = \frac{SS_{\mathrm{Re}gression}}{1} = \frac{1405.2}{1} = 1405.2$$

$$F_0 = \frac{MS_R}{MS_E} = \frac{1405.2}{7.3} = 192.4932$$

P-value is near zero

b) Because the P-value of the F-test ≈ 0 is less than α = 0.05, we reject the null hypothesis that $\beta_1 = 0$ at the 0.05 level of significance. This is the same result obtained from the T_1 test. If the assumptions are valid, a useful linear relationship exists.

c) $\hat{\sigma}^2 = MS_E = 7.3$

11-29 a) 1) The parameter of interest is the regressor variable coefficient, β_1

2) $H_0: \beta_1 = 0$

3) $H_1: \beta_1 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is

$$f_0 = \frac{MS_R}{MS_E} = \frac{SS_R / 1}{SS_E / (n-2)}$$

6) Reject H_0 if $f_0 > f_{\alpha,1,12}$ where $f_{0.05,1,12} = 4.75$
7) Using results from the referenced exercise

$$SS_R = \hat{\beta}_1 S_{xy} = -2.3298017(-59.057143)$$
$$= 137.59$$
$$SS_E = S_{yy} - SS_R$$
$$= 159.71429 - 137.59143$$
$$= 22.123$$
$$f_0 = \frac{137.59}{22.123/12} = 74.63$$

8) Because 74.63 > 4.75 reject H_0 and conclude that compressive strength is significant in predicting intrinsic permeability of concrete at $\alpha = 0.05$. We can therefore conclude that the model specifies a useful linear relationship between these two variables.
$P - \text{value} \cong 0.000002$

b) $\hat{\sigma}^2 = MS_E = \dfrac{SS_E}{n-2} = \dfrac{22.123}{12} = 1.8436$ and $se(\hat{\beta}_1) = \sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\dfrac{1.8436}{25.3486}} = 0.2696$

c) $se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{1.8436\left[\dfrac{1}{14} + \dfrac{3.0714^2}{25.3486}\right]} = 0.9043$

11-30 a) 1) The parameter of interest is the regressor variable coefficient, β_1.
2) $H_0: \beta_1 = 0$
3) $H_1: \beta_1 \neq 0$
4) $\alpha = 0.05$
5) The test statistic is

$$f_0 = \frac{MS_R}{MS_E} = \frac{SS_R/1}{SS_E/(n-2)}$$

6) Reject H_0 if $f_0 > f_{\alpha,1,18}$ where $f_{0.05,1,18} = 4.414$
7) Using the results from Exercise 11-2

$$SS_R = \hat{\beta}_1 S_{xy} = (0.0041612)(141.445) = 0.5886$$
$$SS_E = S_{yy} - SS_R = (8.86 - \tfrac{12.75^2}{20}) - 0.5886 = 0.143275$$
$$f_0 = \frac{0.5886}{0.143275/18} = 73.95$$

8) Because 73.95 > 4.414, reject H_0 and conclude the model specifies a useful relationship at $\alpha = 0.05$.

$$P - value \cong 0.000001$$

b) $se(\hat{\beta}_1) = \sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\dfrac{.00796}{33991.6}} = 4.8391x10^{-4}$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right]} = \sqrt{.00796\left[\frac{1}{20} + \frac{73.9^2}{33991.6}\right]} = 0.04091$$

11-31 a)

Regression Analysis: Rating Pts versus Yds per Att

```
The regression equation is
Rating Pts = 14.2 + 10.1 Yds per Att

Predictor       Coef  SE Coef     T      P
Constant      14.195    9.059  1.57  0.128
Yds per Att   10.092    1.288  7.84  0.000

S = 5.21874   R-Sq = 67.2%   R-Sq(adj) = 66.1%

Analysis of Variance
Source          DF      SS      MS      F      P
Regression       1  1672.5  1672.5  61.41  0.000
Residual Error  30   817.1    27.2
Total           31  2489.5
```

Refer to the ANOVA

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$\alpha = 0.01$$

Because the P-value = 0.000 < α = 0.01, reject H_0. If the assumptions are valid, we conclude that there is a useful linear relationship between these two variables.

b) $\hat{\sigma}^2 = 27.2$

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{27.2}{16.422}} = 1.287$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{27.2\left[\frac{1}{32} + \frac{7^2}{16.422}\right]} = 9.056$$

c) 1)The parameter of interest is the regressor variable coefficient β_1.

2) $H_0{:}\beta_1 = 10$

3) $H_1{:}\beta_1 \neq 10$

4) α = 0.01

5) The test statistic is $t_0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,30} = -2.750$ or $t_0 > t_{0.005,30} = 2.750$

7) Using the results from Exercise 10-6

$$t_0 = \frac{10.092 - 10}{1.287} = 0.0715$$

8) Because 0.0715 < 2.750, fail to reject H_0. There is not enough evidence to conclude that the slope differs from 10 at α = 0.01.

11-32 Refer to ANOVA for the referenced exercise.

a) 1) The parameter of interest is the regressor variable coefficient, β_1.

2) $H_0{:}\beta_1 = 0$

3) $H_1{:}\beta_1 \neq 0$

4) α = 0.05, using t-test

5) The test statistic is $t_0 = \dfrac{\hat{\beta}_1}{se(\hat{\beta}_1)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,22} = -2.074$ or $t_0 > t_{0.025,22} = 2.074$

7) Using the results from Exercise 11-5

$$t_0 = \frac{3.32437}{0.390276} = 8.518$$

8) Since 8.518 > 2.074 reject H_0 and conclude the model is useful $\alpha = 0.05$.

b) 1) The parameter of interest is the slope, β_1

2) $H_0:\beta_1 = 0$

3) $H_1:\beta_1 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is $f_0 = \dfrac{MS_R}{MS_E} = \dfrac{SS_R / 1}{SS_E / (n-2)}$

6) Reject H_0 if $f_0 > f_{\alpha,1,22}$ where $f_{0.01,1,22} = 4.303$

7) Using the results from the referenced exercise

$$f_0 = \frac{636.15569 / 1}{192.89056 / 22} = 72.5563$$

8) Because 72.5563 > 4.303, reject H_0 and conclude the model is useful at a significance $\alpha = 0.05$.

The F-statistic is the square of the t-statistic. The F-test is a restricted to a two-sided test, whereas the t-test could be used for one-sided alternative hypotheses.

c) $se(\hat{\beta}_1) = \sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\dfrac{8.7675}{57.5631}} = .39027$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right]} = \sqrt{8.7675\left[\frac{1}{24} + \frac{6.4049^2}{57.5631}\right]} = 2.5717$$

d) 1) The parameter of interest is the intercept, β_0.

2) $H_0 : \beta_0 = 0$

3) $H_1 : \beta_0 \neq 0$

4) $\alpha = 0.05$, using t-test

5) The test statistic is $t_0 = \dfrac{\hat{\beta}_0}{se(\hat{\beta}_0)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,22} = -2.074$ or $t_0 > t_{0.025,22} = 2.074$

7) Using the results from the referenced exercise

$$t_0 = \frac{13.3201}{2.5717} = 5.179$$

8) Because 5.179 > 2.074 reject H_0 and conclude the intercept is not zero at $\alpha = 0.05$.

11-33 Refer to the ANOVA for the referenced exercise.

a) 1) The parameter of interest is the regressor variable coefficient, β_1.

2) $H_0: \beta_1 = 0$

3) $H_1: \beta_1 \neq 0$

4) $\alpha = 0.01$

5) The test statistic is $f_0 = \dfrac{MS_R}{MS_E} = \dfrac{SS_R\ /1}{SS_E\ /(n-2)}$

6) Reject H_0 if $f_0 > f_{\alpha,1,22}$ where $f_{0.01,1,10} = 10.049$

7) Using the results from the referenced exercise

$$f_0 = \frac{280583.12/1}{37.746089/10} = 74334.4$$

8) Because 74334.4 > 10.049, reject H_0 and conclude the model is useful $\alpha = 0.01$. P-value < 0.000001

b) se($\hat{\beta}_1$) = 0.0337744, se($\hat{\beta}_0$) = 1.66765

c) 1) The parameter of interest is the regressor variable coefficient, β_1.

2) $H_0: \beta_1 = 10$

3) $H_1: \beta_1 \neq 10$

4) $\alpha = 0.01$

5) The test statistic is $t_0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,10} = -3.17$ or $t_0 > t_{0.005,10} = 3.17$

7) Using the results from the referenced exercise

$$t_0 = \frac{9.21-10}{0.0338} = -23.37$$

8) Because $-23.37 < -3.17$ reject H_0 and conclude the slope is not 10 at $\alpha = 0.01$. P-value ≈ 0.

d) H_0: $\beta_0 = 0$ H_1: $\beta_0 \neq 0$

$$t_0 = \frac{-6.3355-0}{1.66765} = -3.8$$

P-value < 0.005. Reject H_0 and conclude that the intercept should be included in the model.

11-34 Refer to the ANOVA for the referenced exercise.

$$H_0 : \beta_1 = 0; H_1 : \beta_1 \neq 0$$

a) $f_0 = \dfrac{MS_R}{MS_E} = \dfrac{385.18}{14.01} = 27.49$

$$F_{0.01,1,19} = 8.18$$

Reject the null hypothesis and conclude that the slope is not zero. The *P*-value ≈ 0.

b) From the computer output in the referenced exercise

$$se(\hat{\beta}_0) = 2.006, \quad se(\hat{\beta}_1) = 0.007671$$

c)

$H_0: \beta_1 = -0.05; H_1: \beta_1 < -0.05$

$$t_0 = \frac{\hat{\beta}_1 - \hat{\beta}_{1,0}}{se(\hat{\beta}_1)} = \frac{-0.040216 - (-0.05)}{0.007671} = \frac{0.090216}{0.007671} = 11.76$$

$t_{0.01,19} = 2.539,$ since t_0 is not less than -$t_{0.01,19} = -2.539,$ do not reject H_0

$P \cong 1.0$

d)

$H_0: \beta_0 = 0; H_1: \beta_0 \neq 0$

$$t_0 = \frac{\hat{\beta}_0 - \hat{\beta}_{0,0}}{se(\hat{\beta}_0)} = \frac{39.156 - 0}{2.006} = 19.52$$

$t_{0.005,19} = 2.861,$ since $|t_0| > t_{0.005,19}$ reject H_0

$P = 4.95E - 14 \cong 0$

11-35 Refer to the ANOVA for the referenced exercise.

a) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 44.0279$

$f_{0.05,1,11} = 4.84$

$f_0 > f_{0.05,1,11}$

Therefore, reject H_0. P-value ≈ 0

b) $se(\hat{\beta}_1) = 0.0104524$

$se(\hat{\beta}_0) = 9.84346$

c) $H_0: \beta_0 = 0$

$H_1: \beta_0 \neq 0$

$\alpha = 0.05$

$t_0 = -1.67718$

$t_{.025,11} = 2.201$

$|t_0| \not< -t_{\alpha/2,11}$

Therefore, fail to reject H_0. P-value = 0.122

11-36 Refer to the ANOVA for the referenced exercise

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.05$

$f_0 = 53.50$

$f_{0.05,1,18} = 4.414$

$f_0 > f_{\alpha,1,18}$

Therefore, reject H_0. P-value ≈ 0

b) $se(\hat{\beta}_1) = 0.0256613$

$se(\hat{\beta}_0) = 2.13526$

c) $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$\alpha = 0.05$

$t_0 = -5.079$

$t_{.025,18} = 2.101$

$| t_0 | > t_{\alpha/2,18}$

Therefore, reject H_0. P-value ≈ 0

11-37 Refer to ANOVA for the referenced exercise

a) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$\alpha = 0.01$

$f_0 = 155.2$

$f_{.01,1,18} = 8.285$

$f_0 > f_{\alpha,1,18}$

Therefore, reject H_0. P-value < 0.00001

b) $se(\hat{\beta}_1) = 45.3468$

$se(\hat{\beta}_0) = 2.96681$

c) $H_0 : \beta_1 = -30$

$H_1 : \beta_1 \neq -30$

$\alpha = 0.01$

$$t_0 = \frac{-36.9618 - (-30)}{2.96681} = -2.3466$$

$t_{.005,18} = 2.878$

$| t_0 | \not> -t_{\alpha/2,18}$

Therefore, fail to reject H_0. P-value = 0.0153(2) = 0.0306

d) $H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

$\alpha = 0.01$

$t_0 = 57.8957$

$t_{0.005,18} = 2.878$

$t_0 > t_{\alpha/2,18}$, therefore, reject H_0. P-value < 0.00001

e) $H_0: \beta_0 = 2500$

$H_1: \beta_0 > 2500$

$\alpha = 0.01$

$$t_0 = \frac{2625.39 - 2500}{45.3468} = 2.7651$$

$t_{.01,18} = 2.552$

$t_0 > t_{\alpha,18}$, therefore reject H_0. P-value = 0.0064

11-38 Refer to the ANOVA for the referenced exercise

a) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$\alpha = 0.01$

$f_0 = 92.224$

$f_{.01,1,16} = 8.531$

$f_0 > f_{\alpha,1,16}$

Therefore, reject H_0.

b) P-value < 0.00001

c) $se(\hat{\beta}_1) = 2.14169$

$se(\hat{\beta}_0) = 1.93591$

d) $H_0: \beta_0 = 0$

$H_1: \beta_0 \neq 0$

$\alpha = 0.01$

$t_0 = 0.243$

$t_{.005,16} = 2.921$

$t_0 \not> t_{\alpha/2,16}$

Therefore, fail to reject H_0. There is not sufficient evidence to conclude that the intercept differs from zero. Based on this test result, the intercept could be removed from the model.

11-39 a) Refer to the ANOVA from the referenced exercise.

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$\alpha = 0.01$$

Because the P-value = 0.000 < α = 0.01, reject H_0. There is evidence of a linear relationship between these two variables.

b) $\hat{\sigma}^2 = 0.083$

The standard errors for the parameters can be obtained from the computer output or calculated as follows.

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{0.083}{420.91}} = 0.014$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{0.083\left[\frac{1}{11} + \frac{10.09^2}{420.91}\right]} = 0.1657$$

c)

1) The parameter of interest is the intercept β_0.

2) $H_0 : \beta_0 = 0$

3) $H_1 : \beta_0 \neq 0$

4) $\alpha = 0.01$

5) The test statistic is $t_0 = \dfrac{\beta_0}{se(\beta_0)}$

6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.005,9} = -3.250$ or $t_0 > t_{\alpha/2,n-2}$ where $t_{0.005,9} = 3.250$

7) Using the results from the referenced exercise

$$t_0 = \frac{0.6578}{0.1657} = 3.97$$

8) Because $t_0 = 3.97 > 3.250$ reject H_0 and conclude the intercept is not zero at α = 0.01.

11-40 a) Refer to the ANOVA for the referenced exercise.

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$\alpha = 0.01$$

Because the P-value = 0.000 < α = 0.01, reject H_0. There is evidence of a linear relationship between these two variables.

b) Yes

c) $\hat{\sigma}^2 = 1.118$

$$se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{1.118}{588}} = 0.0436$$

$$se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{1.118\left[\frac{1}{9} + \frac{65.67^2}{588}\right]} = 2.885$$

11-41 a)

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$\alpha = 0.05$$

Because the P-value = 0.310 > α = 0.05, fail to reject H_0. There is not sufficient evidence of a linear relationship between these two variables.

```
The regression equation is
BMI = 13.8 + 0.256 Age

Predictor    Coef  SE Coef     T      P
Constant   13.820    9.141  1.51  0.174
Age        0.2558   0.2340  1.09  0.310

S = 5.53982   R-Sq = 14.6%   R-Sq(adj) = 2.4%

Analysis of Variance
Source          DF      SS     MS     F      P
Regression       1   36.68  36.68  1.20  0.310
Residual Error   7  214.83  30.69
Total            8  251.51
```

b) $\hat{\sigma}^2 = 30.69, se(\hat{\beta}_1) = 0.2340, se(\hat{\beta}_0) = 9.141$ from the computer output

c) $se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} = \sqrt{30.69\left[\frac{1}{9} + \frac{38.256^2}{560.342}\right]} = 9.141$

11-42 $t_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}}$ After the transformation $\hat{\beta}_1^* = \frac{b}{a}\hat{\beta}_1$, $S_{xx}^* = a^2 S_{xx}$, $\bar{x}^* = a\bar{x}$, $\hat{\beta}_0^* = b\hat{\beta}_0$, and

$\hat{\sigma}^* = b\hat{\sigma}$. Therefore, $t_0^* = \frac{b\hat{\beta}_1 / a}{\sqrt{(b\hat{\sigma})^2 / a^2 S_{xx}}} = t_0$.

11-43 $d = \frac{|10-(12.5)|}{5.5\sqrt{31/16.422}} = 0.331$

Assume α = 0.05, from Chart VII*e* and interpolating between the curves for n = 30 and n = 40, $\beta \cong 0.55$

11-44 a) $\frac{\hat{\beta}}{\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}}}$ has a t distribution with n-1 degree of freedom.

b) From the referenced exercise

, $\hat{\beta} = 21.031, \hat{\sigma} = 3.612,$ and $\sum x_i^2 = 14.707$

The t-statistic in part (a) is 22.331 and $H_0 : \beta_0 = 0$ is rejected for usual α values.

Sections 11-5 and 11-6

11-45

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{5147.996}{2789.282} = 1.846$$

$$\hat{\beta}_0 = 19.032 - (1.846)(25.289) = -27.643$$

$$\hat{y} = -27.643 + 1.846x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{17128.82 - 1.846(5147.996)}{248} = 30.756$$

and $\hat{\sigma} = 5.546$

$$se\hat{\beta}_1 = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{30.756}{2789.282}} = 0.105$$

a) 95% confidence interval on β_1

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$

$$1.846 \pm t_{0.025,248}(0.105)$$

$$1.846 \pm 1.97(0.105)$$

$$1.639 \le \beta_1 \le 2.052$$

b) 95% confidence interval for the mean when BMI is 25

$$\hat{\mu} = -27.643 + 1.846(25) = 18.498$$

$$\hat{\mu} \pm t_{0.025,248}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$18.498 \pm 1.97\sqrt{30.756\left(\frac{1}{250} + \frac{(25 - 25.289)^2}{2789.282}\right)}$$

$$17.805 \le \mu \le 19.191$$

c) 95% prediction interval on $x_0 = 25.0$

$$\hat{y} \pm t_{0.025,248}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$18.498 \pm 1.97\sqrt{30.756\left(1 + \frac{1}{250} + \frac{(25 - 25.289)^2}{2789.282}\right)}$$

$$7.553 \le \mu \le 29.443$$

d) The prediction interval is wider because it includes the variability from a measurement at x_0.

11-46

a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 543503 - \tfrac{11211^2}{250} = 40756.92$$

$$S_{xy} = 1996904.15 - \tfrac{(611211)(44520.8)}{250} = 413.395$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{413.395}{40756.92} = 0.01014$$

$$\hat{\beta}_0 = 178.083 - (0.01014)(44.844) = 177.628$$

$$\hat{y} = 177.628 + 0.01014x$$

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{181998.489 - 0.01014(413.395)}{248} = 733.848$$

$$\hat{\sigma} = 27.090$$

$$se\hat{\beta}_1 = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{27.090}{40756.916}} = 0.134$$

a) 95% confidence interval on β_1

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$

$$0.01014 \pm t_{0.025,248}(0.134)$$

$$0.01014 \pm 1.97(0.134)$$

$$-0.254 \le \beta_1 \le 0.274$$

b) 95% confidence interval for the mean when age is 25

$$\hat{\mu} = 178.083 - (0.01014)(25) = 177.882$$

$$\hat{\mu} \pm t_{0.025,248}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$178.083 \pm 1.97\sqrt{733.848\left(\frac{1}{250} + \frac{(25 - 44.844)^2}{40756.916}\right)}$$

$$171.646 \le \mu \le 184.118$$

c) 95% prediction interval on $x_0 = 25.0$

$$\hat{y} \pm t_{0.025,248}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$

$$177.882 \pm 1.97\sqrt{733.848\left(1 + \frac{1}{250} + \frac{(25 - 44.844)^2}{40756.916}\right)}$$

$$124.164 \le \mu \le 231.600$$

d) The prediction interval is wider because it includes the variability from a measurement at x_0.

e) $s^2 = 730.918$, $\bar{y} = 178.083$

95% confidence interval is
$\bar{y} \pm t_{0.025,248} s/\sqrt{n}$
$178.083 \pm 1.97(\sqrt{730.918/250}$
(174.715, 181.451)

This interval is for the mean weight of all men, not for the specific case of age 25 men. This confidence interval is centered at the mean weight of all men (178.083), so it is centered at greater value than the confidence interval for men of age 25 (177.882). The standard deviation here is estimated from the sample standard deviation of all weights. One might expect that the sample standard deviation (730.198) would be greater than the estimate of σ from the residuals (733.848), but that is not the case here. Consequently, the confidence interval for the age 25 men is unexpectedly wider than the confidence interval for all men.

11-47 $t_{\alpha/2,n-2} = t_{0.025,12} = 2.179$

a) 95% confidence interval on β_1.

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$-2.3298 \pm t_{.025,12}(0.2696)$$
$$-2.3298 \pm 2.179(0.2696)$$
$$-2.9173. \le \beta_1 \le -1.7423.$$

b) 95% confidence interval on β_0.

$$\hat{\beta}_0 \pm t_{.025,12} se(\hat{\beta}_0)$$
$$48.0130 \pm 2.179(0.5959)$$
$$46.7145 \le \beta_0 \le 49.3115.$$

c) 95% confidence interval on μ when $x_0 = 2.5$

$$\hat{\mu}_{Y|x_0} = 48.0130 - 2.3298(2.5) = 42.1885$$
$$\hat{\mu}_{Y|x_0} \pm t_{.025,12}\sqrt{\hat{\sigma}^2\left(\tfrac{1}{n} + \tfrac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$
$$42.1885 \pm (2.179)\sqrt{1.844\left(\tfrac{1}{14} + \tfrac{(2.5-3.0714)^2}{25.3486}\right)}$$
$$42.1885 \pm 2.179(0.3943)$$
$$41.3293 \le \hat{\mu}_{Y|x_0} \le 43.0477$$

d) 95% on prediction interval when $x_0 = 2.5$

$$\hat{y}_0 \pm t_{.025,12}\sqrt{\hat{\sigma}^2\left(1 + \tfrac{1}{n} + \tfrac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$
$$42.1885 \pm 2.179\sqrt{1.844\left(1 + \tfrac{1}{14} + \tfrac{(2.5-3.0714)^2}{25.348571}\right)}$$
$$42.1885 \pm 2.179(1.4056)$$
$$39.1257 \le y_0 \le 45.2513$$

The prediction interval is wider because it includes the variability from a measurement at x_0.

11-48 $t_{\alpha/2,n-2} = t_{0.005,18} = 2.878$

a) $\hat{\beta}_1 \pm \left(t_{0.005,18}\right) se(\hat{\beta}_1)$

$$0.0041612 \pm (2.878)(0.000484)$$
$$0.0027682 \le \beta_1 \le 0.0055542$$

b) $\hat{\beta}_0 \pm \left(t_{0.005,18}\right) se(\hat{\beta}_0)$

$0.3299892 \pm (2.878)(0.04095)$
$0.212250 \le \beta_0 \le 0.447728$

c) 99% confidence interval on μ when $x_0 = 85^\circ$ F

$\hat{\mu}_{Y|x_0} = 0.683689$

$\hat{\mu}_{Y|x_0} \pm t_{.005,18}\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})}$

$0.683689 \pm (2.878)\sqrt{0.00796(\frac{1}{20} + \frac{(85-73.9)^2}{33991.6})}$

0.683689 ± 0.0594607

$0.6242283 \le \hat{\mu}_{Y|x_0} \le 0.7431497$

d) 99% prediction interval when $x_0 = 90^\circ F$.

$\hat{y}_0 = 0.7044949$

$\hat{y}_0 \pm t_{.005,18}\sqrt{\hat{\sigma}^2(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}})}$

$0.7044949 \pm 2.878\sqrt{0.00796(1 + \frac{1}{20} + \frac{(90-73.9)^2}{33991.6})}$

0.7044949 ± 0.263567

$0.420122 \le y_0 \le 0.947256$

11-49 $t_{\alpha/2,n-2} = t_{0.025,30} = 2.042$

a) 95% confidence interval on β_1

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2}se(\hat{\beta}_1)$$
$$10.092 \pm t_{0.025,30}(1.287)$$
$$10.092 \pm 2.042(1.287)$$
$$7.464 \le \beta_1 \le 12.720$$

b) 95% confidence interval on β_0

$$\hat{\beta}_0 \pm t_{\alpha/2,n-2}se(\hat{\beta}_0)$$
$$14.195 \pm t_{0.025,30}(9.056)$$
$$14.195 \pm 2.042(9.056)$$
$$-4.297 \le \hat{\beta}_0 \le 32.687$$

c) 95% confidence interval for the mean rating when the average yards per attempt is 8.0

$$\hat{\mu} = 14.195 + 10.092(8.0) = 94.931$$
$$\hat{\mu} \pm t_{0.025,30}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$
$$94.931 \pm 2.042\sqrt{27.2\left(\frac{1}{32} + \frac{(8-7)^2}{16.422}\right)}$$
$$91.698 \le \mu \le 98.164$$

d) 95% prediction interval on $x_0 = 8.0$

$$\hat{y} \pm t_{0.025,30}\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

$$94.931 \pm 2.042\sqrt{27.2\left(1+\frac{1}{32}+\frac{(8-7)^2}{16.422}\right)}$$

$$83.801 \le \mu \le 106.061$$

11-50

Regression Analysis: Price versus Taxes

```
The regression equation is
Price = 13.3 + 3.32 Taxes


Predictor    Coef  SE Coef     T      P
Constant   13.320    2.572  5.18  0.000
Taxes      3.3244   0.3903  8.52  0.000


S = 2.96104   R-Sq = 76.7%   R-Sq(adj) = 75.7%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  636.16  636.16  72.56  0.000
Residual Error  22  192.89    8.77
Total           23  829.05
```

a) 3.32437 – 2.074(0.3903) = 2.515 ≤ β_1 ≤ 3.32437 + 2.074(0.3903) = 4.134

b) 13.320 – 2.074(0.3903) = 7.985 ≤ β_0 ≤ 13.320 + 2.074(0.39028) = 18.655

c) $38.253 \pm (2.074)\sqrt{8.76775(\frac{1}{24}+\frac{(7.5-6.40492)^2}{57.563139})}$

$$38.253 \pm 1.5353$$

$$36.7177 \le \hat{\mu}_{Y|x_0} \le 39.7883$$

d) $38.253 \pm (2.074)\sqrt{8.76775(1+\frac{1}{24}+\frac{(7.5-6.40492)^2}{57.563139})}$

$$38.253 \pm 6.3302$$

$$31.9228 \le y_0 \le 44.5832$$

11-51

Regression Analysis: Usage versus Temperature

```
The regression equation is
Usage = - 6.34 + 9.21 Temperature
```

```
Predictor       Coef   SE Coef       T      P
Constant      -6.336    1.668   -3.80  0.003
Temperature  9.20836  0.03377  272.64  0.000
S = 1.94284   R-Sq = 100.0%   R-Sq(adj) = 100.0%

Analysis of Variance

Source          DF      SS      MS          F      P
Regression       1  280583  280583   74334.36  0.000
Residual Error  10      38       4
Total           11  280621
```

a) $9.20836 - 2.228(0.03377) = 9.101 \le \beta_1 \le 9.20836 + 2.228(0.03377) = 9.932$

b) $-6.33550 - 2.228(1.66765) = -11.622 \le \beta_0 \le -6.33550 + 2.228(1.66765) = -1.050$

c) $500.124 \pm (2.228)\sqrt{3.774609(\frac{1}{12} + \frac{(55-46.5)^2}{3308.9994})}$

500.124 ± 1.4037586

$498.72024 \le \hat{\mu}_{Y|x_0} \le 501.52776$

d) $500.124 \pm (2.228)\sqrt{3.774609(1 + \frac{1}{12} + \frac{(55-46.5)^2}{3308.9994})}$

500.124 ± 4.5505644

$495.57344 \le y_0 \le 504.67456$

The prediction interval is wider because it includes the variability from a measurement at x_0.

11-52 Refer to the ANOVA for the referenced exercise.

a) $t_{0.025,\,19} = 2.093$

$$34.96 \le \beta_0 \le 43.36; -0.0563 \le \beta_1 \le -0.0241$$

b) Descriptive Statistics: x = displacement

Variable	*n*	Mean	Sum	Sum of Squares
x	21	238.9	5017.0	1436737.0

$$\hat{y} = 33.15 \text{ when } x = 150$$

$$33.15 \pm 2.093\sqrt{14.01\left[\frac{1}{21} + \frac{(150-238.9)^2}{1,436,737.0}\right]}$$

$$33.15 \pm 1.8056$$

$$31.34 \le \mu_{Y|x=150} \le 34.96$$

c)

$$\hat{y} = 33.15 \text{ when } x = 150$$

$$33.15 \pm 2.093\sqrt{14.01\left[1 + \frac{1}{21} + \frac{(150-238.9)^2}{1,436,737.0}\right]}$$

$$33.15 \pm 8.0394$$

$$25.11 \le Y_0 \le 41.19$$

11-53 a) $0.03689 \le \beta_1 \le 0.10183$

b) $-47.0877 \le \beta_0 \le 14.0691$

c) $46.6041 \pm (3.106)\sqrt{7.324951(\frac{1}{13} + \frac{(910-939)^2}{67045.97})}$
46.6041 ± 2.514401
$44.0897 \le \mu_{y|x_0} \le 49.1185)$

d) $46.6041 \pm (3.106)\sqrt{7.324951(1+\frac{1}{13} + \frac{(910-939)^2}{67045.97})}$
46.6041 ± 8.779266
$37.8298 \le y_0 \le 55.3784$

11-54 a) $0.11756 \le \beta_1 \le 0.22541$
b) $-14.3002 \le \beta_0 \le -5.32598$
c) $4.76301 \pm (2.101)\sqrt{1.982231(\frac{1}{20} + \frac{(85-82.3)^2}{3010.2111})}$
4.76301 ± 0.6772655
$4.0857 \le \mu_{y|x_0} \le 5.4403$
d) $4.76301 \pm (2.101)\sqrt{1.982231(1+\frac{1}{20} + \frac{(85-82.3)^2}{3010.2111})}$
4.76301 ± 3.0345765
$1.7284 \le y_0 \le 7.7976$

11-55 a) $201.552 \le \beta_1 \le 266.590$
b) $-4.67015 \le \beta_0 \le -2.34696$
c) $128.814 \pm (2.365)\sqrt{398.2804(\frac{1}{9} + \frac{(30-24.5)^2}{1651.4214})}$
128.814 ± 16.980124
$111.8339 \le \mu_{y|x_0} \le 145.7941$

11-56 a) $14.3107 \le \beta_1 \le 26.8239$
b) $-5.18501 \le \beta_0 \le 6.12594$
c) $21.038 \pm (2.921)\sqrt{13.8092(\frac{1}{18} + \frac{(1-0.806111)^2}{3.01062})}$
21.038 ± 2.8314277
$18.2066 \le \mu_{y|x_0} \le 23.8694$
d) $21.038 \pm (2.921)\sqrt{13.8092(1+\frac{1}{18} + \frac{(1-0.806111)^2}{3.01062})}$
21.038 ± 11.217861
$9.8201 \le y_0 \le 32.2559$

11-57 a) $-43.1964 \le \beta_1 \le -30.7272$
b) $2530.09 \le \beta_0 \le 2720.68$
c) $1886.154 \pm (2.101)\sqrt{9811.21(\frac{1}{20} + \frac{(20-13.3375)^2}{1114.6618})}$
1886.154 ± 62.370688
$1823.7833 \le \mu_{y|x_0} \le 1948.5247$
d) $1886.154 \pm (2.101)\sqrt{9811.21(1+\frac{1}{20} + \frac{(20-13.3375)^2}{1114.6618})}$
1886.154 ± 217.25275
$1668.9013 \le y_0 \le 2103.4067$

11-58 $t_{\alpha/2,n-2} = t_{0.025,5} = 2.571$

a) 95% confidence interval on $\hat{\beta}_1$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$-0.034 \pm t_{0.025,5}(0.026)$$
$$-0.034 \pm 2.571(0.026)$$
$$-0.101 \leq \hat{\beta}_1 \leq 0.033$$

b) 95% confidence interval on β_0

$$\hat{\beta}_0 \pm t_{\alpha/2,n-2} se(\hat{\beta}_0)$$
$$55.63 \pm t_{0.025,5}(32.11)$$
$$55.63 \pm 2.571(32.11)$$
$$-26.89 \leq \hat{\beta}_0 \leq 138.15$$

c) 95% confidence interval for the mean length when x=1500:

$$\hat{\mu} = 55.63 - 0.034(1500) = 4.63$$
$$\hat{\mu} \pm t_{0.025,5}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$
$$4.63 \pm 2.571\sqrt{77.33\left(\frac{1}{7} + \frac{(1500 - 1242.86)^2}{117142.8}\right)}$$
$$4.63 \pm 2.571(7.396)$$
$$-14.39 \leq \mu \leq 4.63$$

d) 95% prediction interval when $x_0 = 1500$

$$\hat{y} \pm t_{0.025,5}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$
$$4.63 \pm 2.571\sqrt{77.33\left(1 + \frac{1}{7} + \frac{(1500 - 1242.86)^2}{117142.8}\right)}$$
$$4.63 \pm 2.571(11.49)$$
$$-24.91 \leq y_0 \leq 34.17$$

The prediction interval is wider because it includes the variability from a measurement at x_0.

11-59 Refer to the computer output in the referenced exercise.

$$t_{\alpha/2,n-2} = t_{0.005,9} = 3.250$$

a) 99% confidence interval for $\hat{\beta}_1$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$0.178 \pm t_{0.005,9}(0.014)$$
$$0.178 \pm 3.250(0.014)$$
$$0.1325 \leq \hat{\beta}_1 \leq 0.2235$$

b) 99% confidence interval on β_0

$$\hat{\beta}_0 \pm t_{\alpha/2,n-2} se(\hat{\beta}_0)$$
$$0.6578 \pm t_{0.005,9}(0.1657)$$
$$0.6578 \pm 3.250(0.1657)$$
$$0.119 \leq \hat{\beta}_0 \leq 1.196$$

c) 95% confidence interval on μ when $x_0 = 8$

$$\hat{\mu}_{y|x_0} = 0.658 + 0.178(8) = 2.082$$
$$\hat{\mu}_{y|x_0} \pm t_{0.025,9}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$$
$$2.082 \pm 2.262\sqrt{0.083\left(\frac{1}{11} + \frac{(8 - 10.09)^2}{420.91}\right)}$$
$$1.87 \leq \mu_{y|x_0} \leq 2.29$$

Section 11-7

11-60

Results depend on the simulated errors. Example data follow for the model y = 10 + 30x.

Model y = 10 + 30x

x	e	y
10	-1.50427	308.4957
12	-0.80622	369.1938
14	1.836897	431.8369
16	0.35073	490.3507
18	0.80104	550.801
20	-0.26163	609.7384
22	-0.70788	669.2921
24	0.36672	730.3667

a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 2480 - \frac{136^2}{8} = 168$$

$$S_{xy} = 75769.025 - \frac{(136)(4160.075)}{8} = 5047.743$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{5047.743}{168} = 30.046$$

$$\hat{\beta}_0 = 520.009 - (30.046)(17) = 9.226$$

$$\hat{y} = 9.226 + 30.046x$$

b) $\hat{\sigma}^2 = MS_E = \dfrac{SS_E}{n-2} = \dfrac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \dfrac{151672.357 - 30.046(5047.743)}{6} = 1.233$

$\hat{\sigma} = 1.110$

c)

$$SS_R = \hat{\beta}_1 S_{xy} = 30.046(5047.743) = 151664.958$$

$$SS_T = S_{yy} = 151672.357$$

$$R^2 = \frac{SS_R}{SS_T} = \frac{151664.958}{151672.357} = 0.99995$$

d) Results will depend on the simulated errors. Example data follow.

x	e	y
10	-1.50427	308.4957
14	-0.80622	429.1938
18	1.836897	551.8369
22	0.35073	670.3507
26	0.80104	790.801
30	-0.26163	909.7384
34	-0.70788	1029.292
38	0.36672	1150.367

d) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$S_{xx} = 5280 - \tfrac{192^2}{8} = 672$$
$$S_{xy} = 160337.296 - \tfrac{(192)(5840.025)}{8} = 20175.487$$
$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{20175.487}{672} = 30.023$$
$$\hat{\beta}_0 = 730.009 - (30.023)(24) = 9.456$$
$$\hat{y} = 9.456 + 30.023x$$
$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{605736.958 - 30.023(20175.487)}{6} = 1.233$$
$$\hat{\sigma} = 1.110$$

e)
$$SS_R = \hat{\beta}_1 S_{xy} = 30.023(20175.487) = 605729.56$$
$$SS_T = S_{yy} = 605736.958$$
$$R^2 = \frac{SS_R}{SS_T} = 0.999987$$
Because the dispersion of x increased, R^2 increased.

11-61
Results will depend on the simulated errors. Example data follow.

x	e	y
10	-1.01617	308.98383
12	-9.26768	360.73232
14	-5.13715	424.86285
16	-1.71362	488.28638
18	3.776995	553.777
20	-4.28057	605.71943
22	-4.05776	665.94224
24	2.337083	732.33708

a)
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$S_{xx} = 2480 - \tfrac{136^2}{8} = 168$$
$$S_{xy} = 75488.482 - \tfrac{(136)(4140.641)}{8} = 5097.583$$
$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{5087.583}{168} = 30.343$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 517.580 - (30.343)(17) = 1.753$$
$$\hat{y} = 1.753 + 30.343x$$

b) $$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{154778.884 - 30.343(5097.583)}{6} = 17.364$$
$$\hat{\sigma} = 4.167$$

c)

$$SS_R = \hat{\beta}_1 S_{xy} = 30.343(5097.583) = 154674.700$$
$$SS_T = S_{yy} = 154778.884$$
$$R^2 = \frac{SS_R}{SS_T} = 0.9993$$

Notice that R^2 decreased with the increased standard deviation of noise.

d) Results will depend on the simulated errors. Example data follow.

x	e	y
10	-1.01617	308.98383
14	-9.26768	420.73232
18	-5.13715	544.86285
22	-1.71362	668.28638
26	3.776995	793.777
30	-4.28057	905.71943
34	-4.05776	1025.9422
38	2.337083	1152.3371

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$S_{xx} = 5280 - \tfrac{192^2}{8} = 672$$
$$S_{xy} = 159970.553 - \tfrac{(192)(5820,641)}{8} = 20275.165$$
$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{20275.17}{672} = 30.171$$
$$\hat{\beta}_0 = 727.580 - (30.171)(24) = 3.467$$
$$\hat{y} = 3.467 + 30.171x$$
$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = \frac{611833.847 - 30.171(20275.517)}{6} = 17.364$$
$$\hat{\sigma} = 4.167$$

e)

$$SS_R = \hat{\beta}_1 S_{xy} = 30.171(20275.165) = 611729.664$$
$$SS_T = S_{yy} = 611833.847$$
$$R^2 = \frac{SS_R}{SS_T} = \frac{611729.664}{611833.847} = 0.9998$$

Because the dispersion of x increased, R^2 increased.

11-62 $$R^2 = \hat{\beta}_1^2 \frac{S_{XX}}{S_{YY}} = (-2.330)^2 \frac{25.35}{159.71} = 0.8617$$

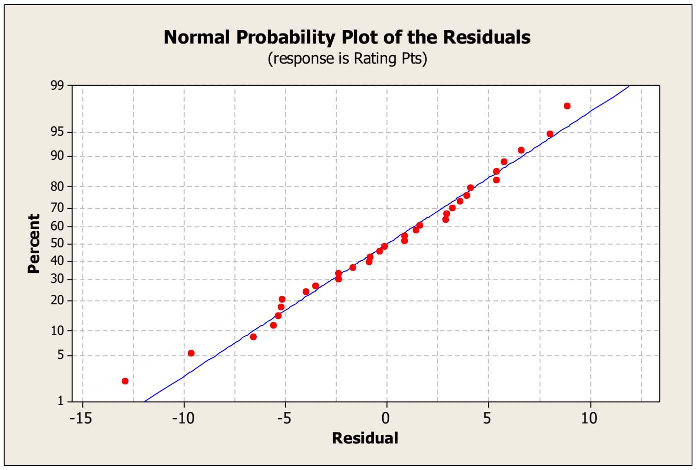
The model accounts for 86.17% of the variability in the data.

11-63 Refer to the computer output in the referenced exercise.

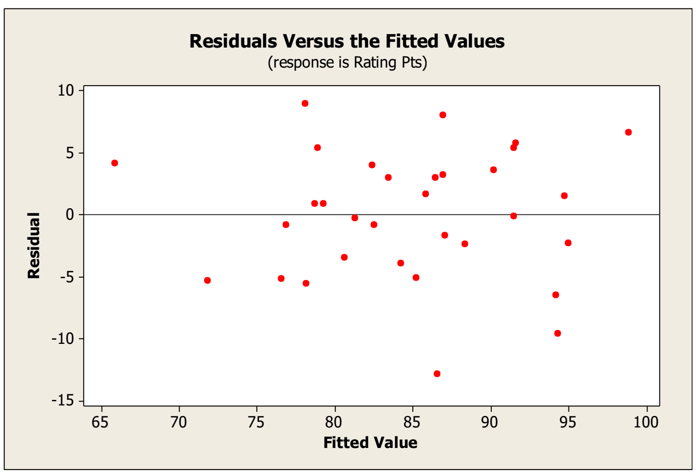
a) $R^2 = 0.672$

The model accounts for 67.2% of the variability in the data.

b) There is no major departure from the normality assumption in the following graph.



c) The assumption of constant variance appears reasonable.



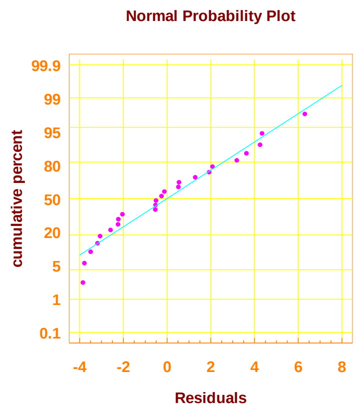
11-64 Use the results from the referenced exercise to answer the following questions.

a)

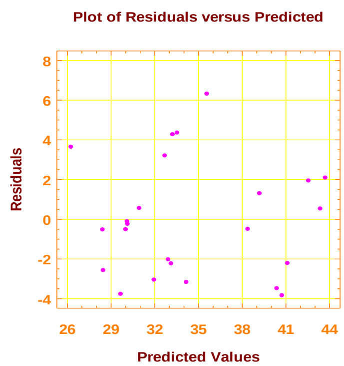
SalePrice	Taxes	Predicted	Residuals
25.9	4.9176	29.6681073	-3.76810726
29.5	5.0208	30.0111824	-0.51118237
27.9	4.5429	28.4224654	-0.52246536
25.9	4.5573	28.4703363	-2.57033630
29.9	5.0597	30.1405004	-0.24050041
29.9	3.8910	26.2553078	3.64469225
30.9	5.8980	32.9273208	-2.02732082
28.9	5.6039	31.9496232	-3.04962324
35.9	5.8282	32.6952797	3.20472030
31.5	5.3003	30.9403441	0.55965587
31.0	6.2712	34.1679762	-3.16797616
30.9	5.9592	33.1307723	-2.23077234
30.0	5.0500	30.1082540	-0.10825401
36.9	8.2464	40.7342742	-3.83427422

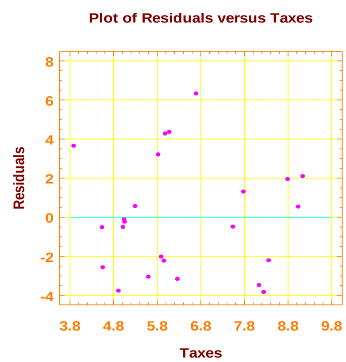
41.9	6.6969	35.5831610	6.31683901
40.5	7.7841	39.1974174	1.30258260
43.9	9.0384	43.3671762	0.53282376
37.5	5.9894	33.2311683	4.26883165
37.9	7.5422	38.3932520	-0.49325200
44.5	8.7951	42.5583567	1.94164328
37.9	6.0831	33.5426619	4.35733807
38.9	8.3607	41.1142499	-2.21424985
36.9	8.1400	40.3805611	-3.48056112
45.8	9.1416	43.7102513	2.08974865

b) Assumption of normality does not seem to be violated because the data fall along a line.



c) There are no serious departures from the assumption of constant variance. This is evident by the random pattern of the residuals.



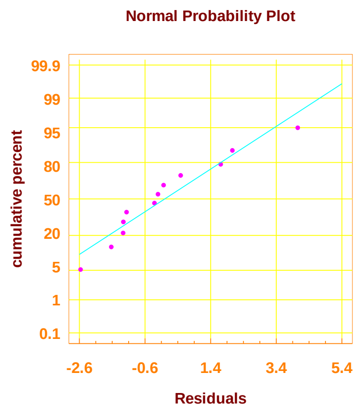


d) $R^2 = 76.73\%$

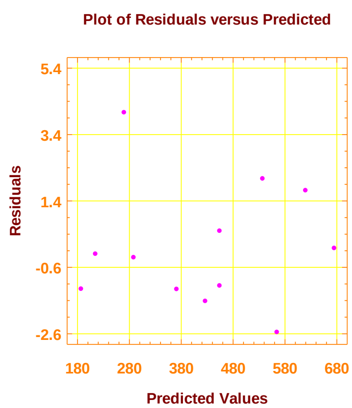
11-65 Use the results of the referenced exercise to answer the following questions

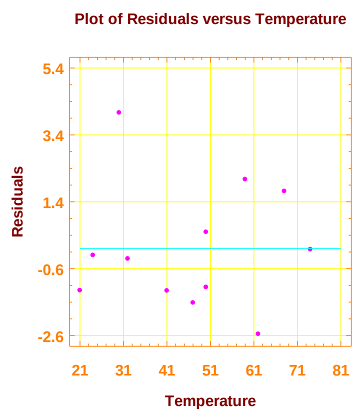
a) $R^2 = 99.986\%$; The proportion of variability explained by the model.

b) Yes, normality seems to be satisfied because the data appear to fall along a line.



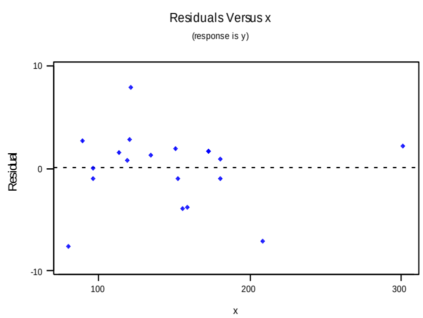
c) There might be lower variance at the middle settings of x. However, this data does not indicate a serious departure from the assumptions.

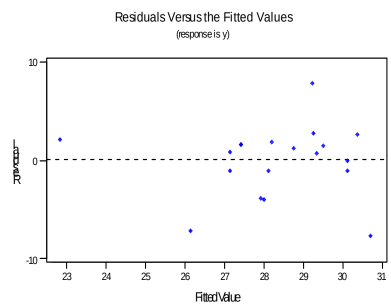




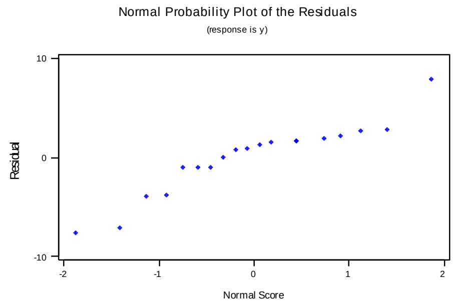
11-66 a) $R^2 = 20.1121\%$

b) These plots might indicate the presence of outliers, but no real problem with assumptions.



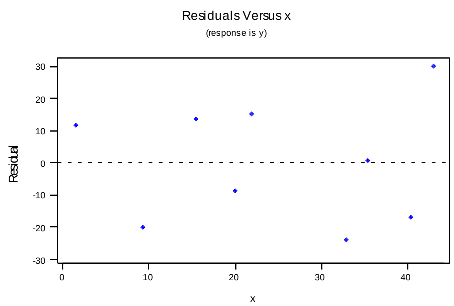


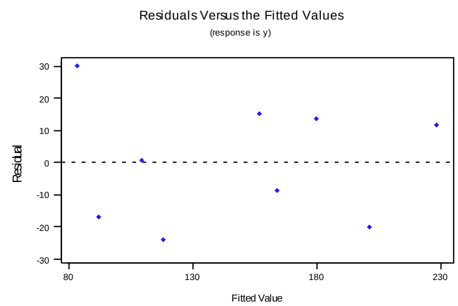
c) The normality assumption appears marginal.



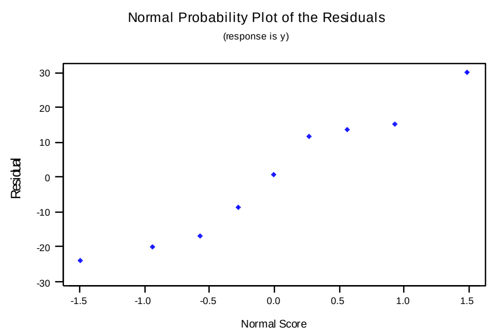
11-67 a) $R^2 = 0.879397$

b) No departures from constant variance are noted.



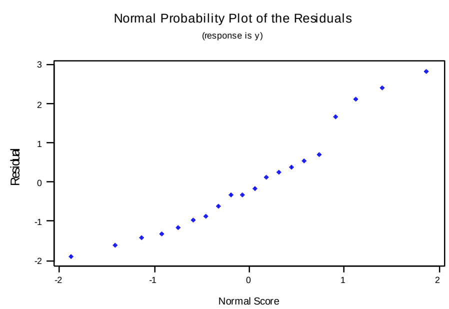


c) Normality assumption appears reasonable.

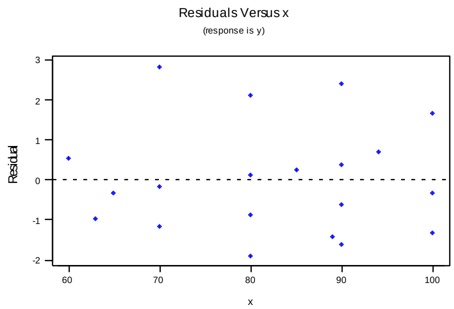


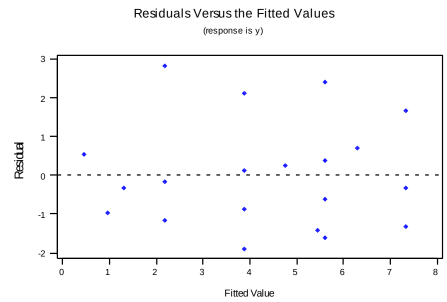
11-68 a) $R^2 = 71.27\%$

b) No major departure from normality assumptions.



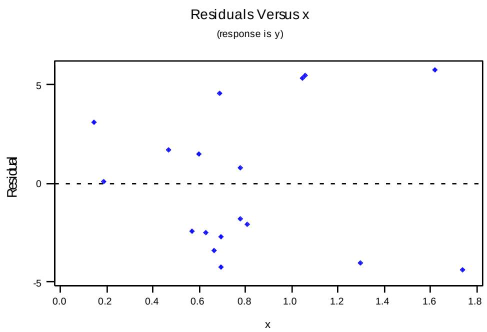
c) Assumption of constant variance appears reasonable.

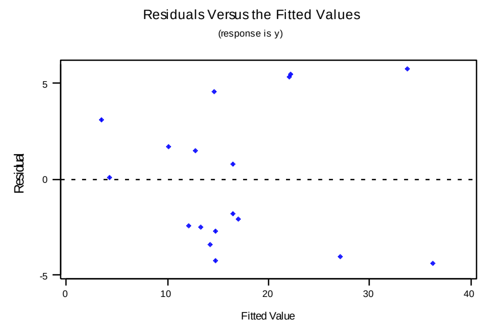




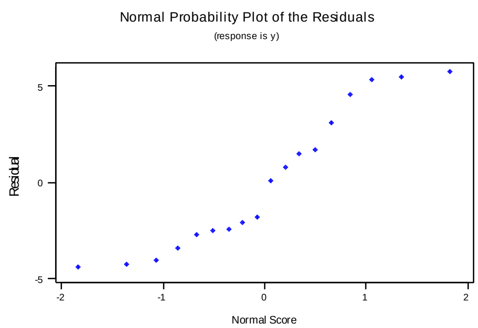
11-69 a) $R^2 = 85.22\%$

b) Assumptions appear reasonable, but there is a suggestion that variability increases slightly with $\hat{y}$.





c) Normality assumption may be questionable. There is some "bending" away from a line in the tails of the normal probability plot.



11-70 a)

```
The regression equation is
Compressive Strength = - 2150 + 185 Density

Predictor      Coef  SE Coef      T      P
Constant    -2149.6    332.5  -6.46  0.000
Density      184.55    11.79  15.66  0.000

S = 339.219   R-Sq = 86.0%   R-Sq(adj) = 85.6%

Analysis of Variance

Source          DF        SS        MS       F      P
Regression       1  28209679  28209679  245.15  0.000
Residual Error  40   4602769    115069
Total           41  32812448
```

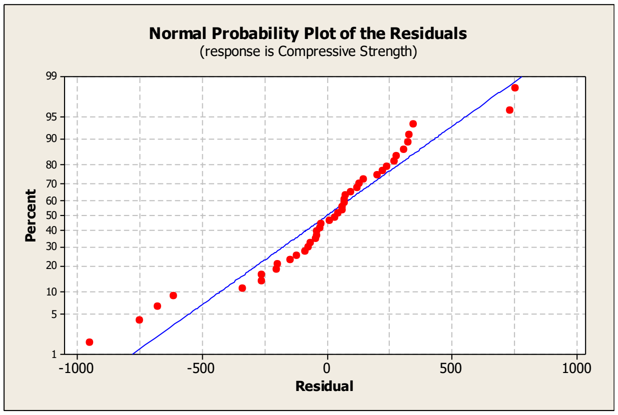
b) Because the P-value = 0.000 < α = 0.05, the model is significant.

c) $\hat{\sigma}^2 = 115069$

d) $R^2 = \dfrac{SS_R}{SS_T} = 1 - \dfrac{SS_E}{SS_T} = \dfrac{28209679}{32812448} = 0.8597 = 85.97\%$

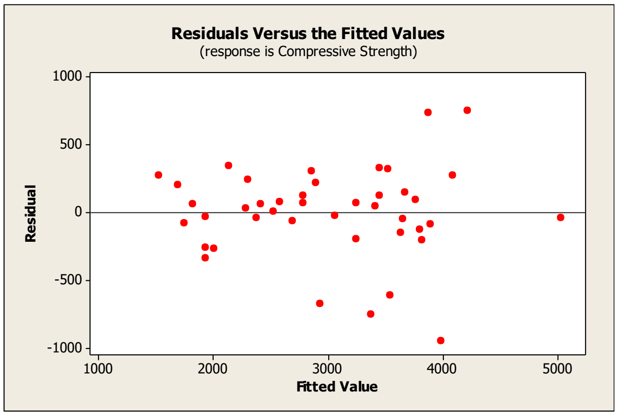
The model accounts for 85.97% of the variability in the data.

e)



No major departure from the normality assumption.

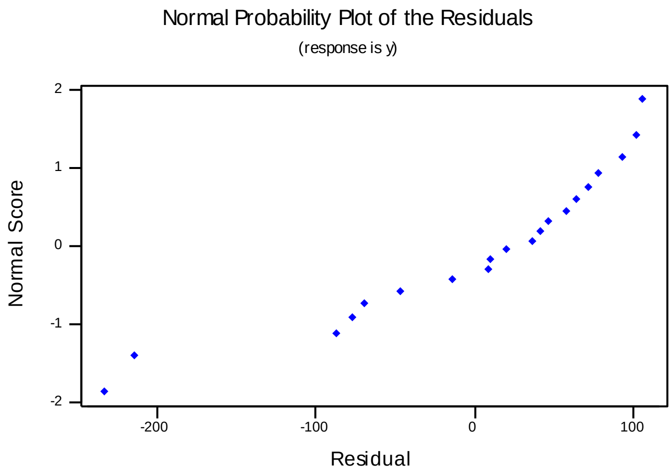
f)



Assumption of constant variance appears reasonable.

11-71 a) $R^2 = 0.896081$ 89% of the variability is explained by the model.

b) Yes, the two points with residuals much larger in magnitude than the others seem unusual.



c) $R^2_{new\ \ model} = 0.9573$

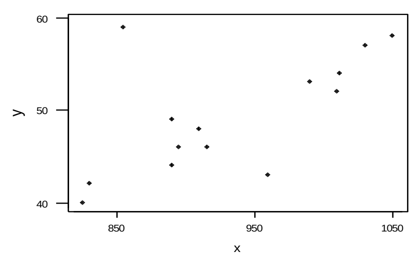
Larger, because the model is better able to account for the variability in the data with these two outlying data points removed.

d) $\hat{\sigma}^2_{old\ \ model} = 9811.21$

$\hat{\sigma}^2_{\text{new model}} = 4022.93$

Yes, reduced more than 50%, because the two removed points accounted for a large amount of the error.

11-72 a)



$\hat{y} = 0.677559 + 0.0521753x$

b) $H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$ $\alpha = 0.05$

$f_0 = 7.9384$

$f_{0.05,1,12} = 4.75$

$f_0 > f_{\alpha,1,12}$

Reject H_0.

c) $\hat{\sigma}^2 = 25.23842$

d) $\hat{\sigma}_{orig}{}^2 = 7.502$

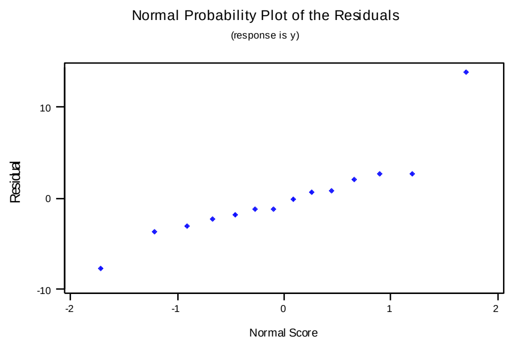
The new estimate is larger because the new point added additional variance that was not accounted for by the model.

e) $\hat{y} = 0.677559 + 0.0521753(855) = 45.287$

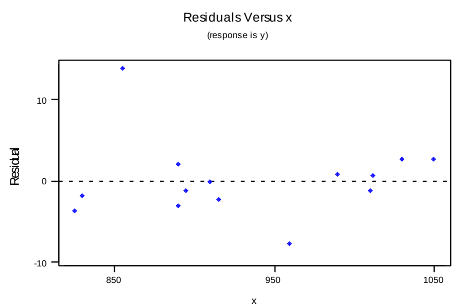
$e = y - \hat{y} = 59 - 45.287 = 13.713$

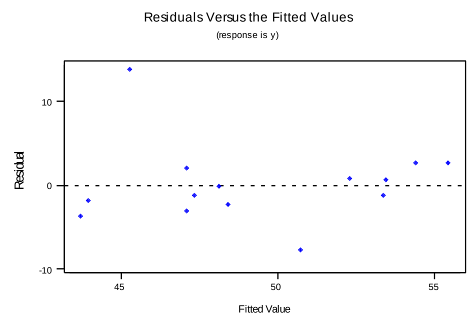
Yes, e_{14} is especially large compared to the other residuals.

f) The one added point is an outlier and the normality assumption is not as valid with the point included.



g) Constant variance assumption appears valid except for the added point.





11-73 Yes, when the residuals are standardized the unusual residuals are easier to identify.

1.0723949	0.205124
-0.70057	-0.88064
-0.13822	0.792107
0.648804	0.72429
-2.35308	-0.47375
-2.16868	0.943215
0.468419	0.108886
0.424214	0.374926
0.098212	1.037219
0.594582	-0.77744

11-74 For two random variables X_1 and X_2,

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2)$$

Then,

$$\begin{aligned} V(Y_i - \hat{Y}_i) &= V(Y_i) + V(\hat{Y}_i) - 2Cov(Y_i, \hat{Y}_i) \\ &= \sigma^2 + V(\hat{\beta}_0 + \hat{\beta}_1 x_i) - 2\sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right] \\ &= \sigma^2 + \sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right] - 2\sigma^2\left[\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right] \\ &= \sigma^2\left[1 - \left(\tfrac{1}{n} + \tfrac{(x_i-\bar{x})^2}{S_{xx}}\right)\right] \end{aligned}$$

a) Because e_i is divided by an estimate of its standard error (when σ^2 is estimated by $\hat{\sigma}^2$), r_i has approximately unit variance.

b) No, the term in brackets in the denominator is necessary.

c) If x_i is near $\bar{X}$ and n is reasonably large, r_i is approximately equal to the standardized residual.

d) If x_i is far from $\bar{X}$, the standard error of e_i is small. Consequently, extreme points are better fit by least squares regression than points near the middle range of x. Because the studentized residual at any point has variance of approximately one, the studentized residuals can be used to compare the fit of points to the regression line over the range of *x*.

11-75 Using $R^2 = 1 - \frac{SS_E}{S_{yy}}$, $F_0 = \dfrac{(n-2)(1-\frac{SS_E}{S_{yy}})}{\frac{SS_E}{S_{yy}}} = \dfrac{S_{yy} - SS_E}{\frac{SS_E}{n-2}} = \dfrac{S_{yy} - SS_E}{\hat{\sigma}^2}$

Also,

$$\begin{aligned}
SS_E &= \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\
&= \sum (y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x}))^2 \\
&= \sum (y_i - \bar{y}) + \hat{\beta}_1{}^2 \sum (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum (y_i - \bar{y})(x_i - \bar{x}) \\
&= \sum (y_i - \bar{y})^2 - \hat{\beta}_1{}^2 \sum (x_i - \bar{x})^2
\end{aligned}$$

$$S_{yy} - SS_E = \hat{\beta}_1{}^2 \sum (x_i - \bar{x})^2$$

Therefore, $F_0 = \dfrac{\hat{\beta}_1^2}{\hat{\sigma}^2 / S_{xx}} = t_0^2$

Because the square of a *t* random variable with *n*-2 degrees of freedom is an *F* random variable with 1 and *n*-2 degrees of freedom, the usually *t*-test that compares $|\, t_0 \,|$ to $t_{\alpha/2,n-2}$ is equivalent to comparing $f_0 = t_0^2$ to $f_{\alpha,1,n-2} = t^2_{\alpha/2,n-2}$.

a) $f_0 = \dfrac{0.9(23)}{1-0.9} = 207$. Reject $H_0 : \beta_1 = 0$.

b) Because $f_{0.05,1,23} = 4.28$, H_0 is rejected if $\dfrac{23R^2}{1-R^2} > 4.28$

That is, H_0 is rejected if

$$\begin{aligned}
23R^2 &> 4.28(1 - R^2) \\
27.28R^2 &> 4.28 \\
R^2 &> 0.157
\end{aligned}$$

Section 11-8

11-76 a) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$ α = 0.05

$t_0 = \frac{0.8\sqrt{20-2}}{\sqrt{1-0.64}} = 5.657$

$t_{0.025,18} = 2.101$

$|t_0| > t_{0.025,18}$

Reject H0. P-value < 2(0.0005) = 0.001

b) $H_0 : \rho = 0.5$

$H_1 : \rho \neq 0.5$ α = 0.05

$z_0 = (\text{arctanh }(0.8) - \text{arctanh }(0.5))(17)^{1/2} = 2.265$

$z_{.025} = 1.96$

$|z_0| > z_{\alpha/2}$

Reject H0. P-value = 2(0.012) = 0.024

c) $\tanh(\text{arctanh } 0.8 - \frac{z_{.025}}{\sqrt{17}}) \leq \rho \leq \tanh(\text{arctanh } 0.8 + \frac{z_{.025}}{\sqrt{17}})$

where $z_{0.025} = 1.96$

$0.5534 \leq \rho \leq 0.9177$

Because ρ = 0 and ρ = 0.5 are not in the interval, reject H0.

11-77 a) $H_0 : \rho = 0$

$H_1 : \rho > 0$ α = 0.05

$t_0 = \frac{0.75\sqrt{20-2}}{\sqrt{1-0.75^2}} = 4.81$

$t_{0.05,18} = 1.734$

$t_0 > t_{0.05,18}$

Reject H0. P-value < 0.0005

b) $H_0 : \rho = 0.5$

$H_1 : \rho > 0.5$ α = 0.05

$z_0 = (\text{arctanh }(0.75) - \text{arctanh }(0.5))(17)^{1/2} = 1.7467$

$z_{.05} = 1.65$

$z_0 > z_\alpha$

Reject H0. P-value = 0.04

c) $\rho \geq \tanh(\text{arctanh } 0.75 - \frac{z_{0.05}}{\sqrt{17}})$ where $z_{.05} = 1.64$

$\rho \geq 2.26$

Because ρ = 0 and ρ = 0.5 are not in the interval, reject the null hypotheses from parts (a) and (b).

11-78 n = 25 r = 0.83

a) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$ α = 0.05

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.83\sqrt{23}}{\sqrt{1-(0.83)^2}} = 7.137$

$t_{.025,23} = 2.069$

$t_0 > t_{\alpha/2,23}$

Reject H_0 , P-value ≈ 0

b) $\tanh(\text{arctanh } 0.83 - \frac{z_{.025}}{\sqrt{22}}) \le \rho \le \tanh(\text{arctanh } 0.83 + \frac{z_{.025}}{\sqrt{22}})$

where $z_{.025} = 196$. $0.6471 \le \rho \le 0.9226$.

c) $H_0 : \rho = 0.8$

$H_1 : \rho \ne 0.8$ α = 0.05

$z_0 = (\text{arctanh } 0.83 - \text{arctanh } 0.8)(22)^{1/2} = 0.4199$

$z_{.025} = 1.96$

$z_0 \not> z_{\alpha/2}$

Do not reject H_0. P-value = (0.3373)(2) = 0.675

11-79 n = 50 r = 0.62

a) $H_0 : \rho = 0$

$H_1 : \rho \ne 0$ α = 0.01

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.62\sqrt{48}}{\sqrt{1-(0.62)^2}} = 5.475$

$t_{.005,48} = 2.682$

$t_0 > t_{0.005,48}$

Reject H_0. *P*-value ≅ 0

b) $\tanh(\text{arctanh } 0.62 - \frac{z_{.005}}{\sqrt{47}}) \le \rho \le \tanh(\text{arctanh } 0.62 + \frac{z_{.005}}{\sqrt{47}})$

where $z_{0.005} = 2.575$.

$0.3358 \le \rho \le 0.8007$

c) Yes.

11-80 a) r = 0.933203

a) $H_0 : \rho = 0$

$H_1 : \rho \ne 0$ α = 0.05

$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.933203\sqrt{15}}{\sqrt{1-(0.8709)}} = 10.06$

$t_{.025,15} = 2.131$

$t_0 > t_{\alpha/2,15}$

Reject H_0

c) $\hat{y} = 0.72538 + 0.498081x$

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \ne 0$ α = 0.05

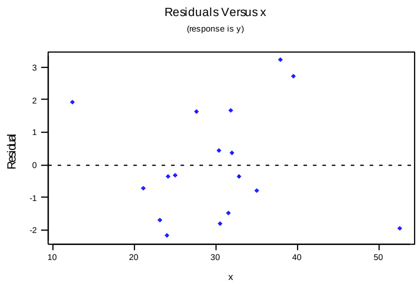
$f_0 = 101.16$

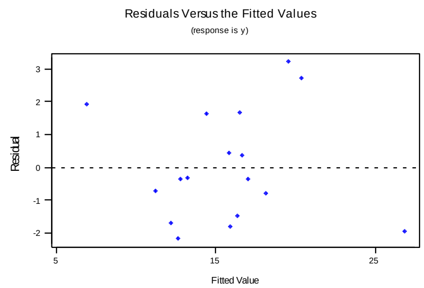
$f_{0.05,1,15} = 4.543$

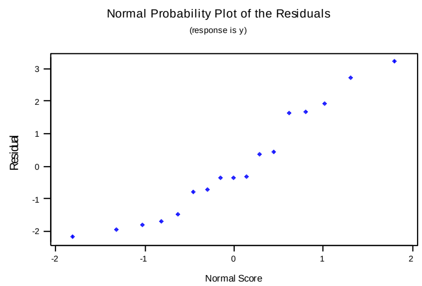
$f_0 >> f_{\alpha,1,15}$

Reject H_0. Conclude that the model is significant at α = 0.05. This test and the one in part b) are identical.

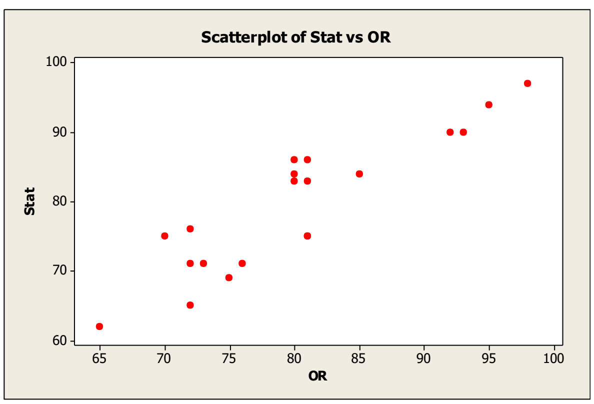
d) No problems with model assumptions are noted.







11-81 a) $\hat{y} = -0.0280411 + 0.990987x$



b) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$ α = 0.05

$f_0 = 79.838$

$f_{0.05,1,18} = 4.41$

$f_0 >> f_{\alpha,1,18}$

Reject H_0

c) $r = \sqrt{0.816} = 0.903$

d) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$ α = 0.05

$$t_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} = \frac{0.90334\sqrt{18}}{\sqrt{1-0.816}} = 8.9345$$

$t_{0.025,18} = 2.101$

$t_0 > t_{\alpha/2,18}$

Reject H_0

e) $H_0 : \rho = 0.5$

$H_1 : \rho \neq 0.5$ α = 0.05

$z_0 = 3.879$

$z_{.025} = 1.96$

$z_0 > z_{\alpha/2}$

Reject H_0

f) $\tanh(\text{arctanh } 0.90334 - \frac{z_{.025}}{\sqrt{17}}) \leq \rho \leq \tanh(\text{arctanh } 0.90334 + \frac{z_{.025}}{\sqrt{17}})$ where $z_{0.025} = 1.96$.

$0.7677 \leq \rho \leq 0.9615$

11-82 a) $\hat{y} = 69.1044 + 0.419415x$

b) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$ α = 0.05

$f_0 = 35.744$

$f_{0.05,1,24} = 4.260$

$f_0 > f_{\alpha,1,24}$

Reject H_0

c) $r = 0.77349$

d) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$ α = 0.05

$t_0 = \frac{0.77349\sqrt{24}}{\sqrt{1-0.5983}} = 5.9787$

$t_{0.025,24} = 2.064$

$t_0 > t_{\alpha/2,24}$

Reject H_0

e) $H_0 : \rho = 0.6$

$H_1 : \rho \neq 0.6$ α = 0.05

$z_0 = (\text{arctanh } 0.77349 - \text{arctanh } 0.6)(23)^{1/2} = 1.6105$

$z_{.025} = 1.96$

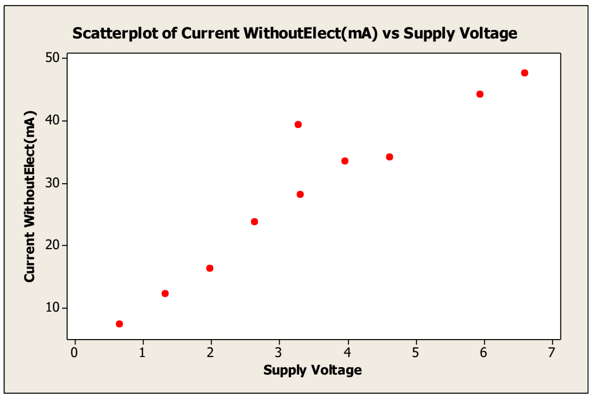
$z_0 \not> z_{\alpha/2}$

Fail to reject H_0

f) $\tanh(\text{arctanh } 0.77349 - \frac{z_{.025}}{\sqrt{23}}) \leq \rho \leq \tanh(\text{arctanh } 0.77349 + \frac{z_{.025}}{\sqrt{23}})$ where $z_{0.025} = 1.96$

$0.5513 \leq \rho \leq 0.8932$

11-83 a)



```
The regression equation is
Current WithoutElect(mA) = 5.50 + 6.73 Supply Voltage


Predictor         Coef  SE Coef     T      P
Constant         5.503    3.104  1.77  0.114
Supply Voltage  6.7342   0.7999  8.42  0.000


S = 4.59061   R-Sq = 89.9%   R-Sq(adj) = 88.6%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       1  1493.7  1493.7  70.88  0.000
Residual Error   8   168.6    21.1
Total            9  1662.3
```

$\hat{y} = 5.50 + 6.73x$

Yes, because the P-value ≈ 0, the regression model is significant at $\alpha = 0.05$.

b) $r = \sqrt{0.899} = 0.948$

c)

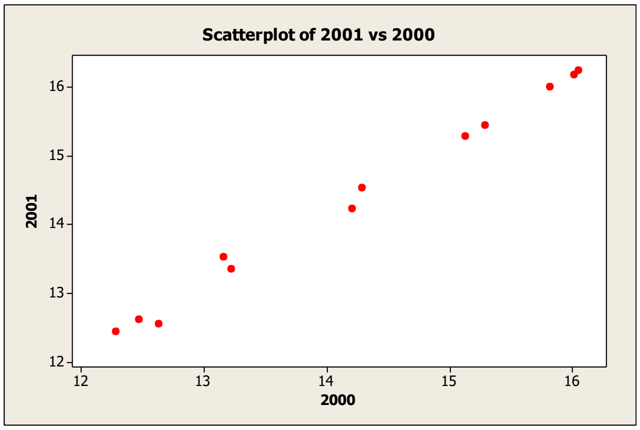
$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$
$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.948\sqrt{10-2}}{\sqrt{1-0.948^2}} = 8.425$$
$$t_{0.025,8} = 2.306$$
$$t_0 = 8.425 > t_{0.025,8} = 2.306$$

Reject H_0

d)

$$\tanh\left(\text{arctan h} \quad r - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \leq \rho \leq \tanh\left(\text{arctan h} \quad r + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right)$$
$$\tanh\left(\text{arctan h} \quad 0.948 - \frac{1.96}{\sqrt{10-3}}\right) \leq \rho \leq \tanh\left(\text{arctan h} \quad 0.948 + \frac{19.6}{\sqrt{10-3}}\right)$$
$$0.7898 \leq \rho \leq 0.9879$$

11-84 a)




```
The regression equation is
Y2001 = - 0.014 + 1.01 Y2000

Predictor      Coef   SE Coef      T      P
Constant    -0.0144    0.3315  -0.04  0.966
Y2000       1.01127   0.02321  43.56  0.000

S = 0.110372   R-Sq = 99.5%   R-Sq(adj) = 99.4%

Analysis of Variance

Source          DF      SS      MS        F      P
Regression       1  23.117  23.117  1897.63  0.000
Residual Error  10   0.122   0.012
Total           11  23.239
```

$$\hat{y} = -0.014 + 1.011x$$

Yes, because the P-value ≈ 0, the regression model is significant at α = 0.05.

b) $r = \sqrt{0.995} = 0.9975$

c)

$H_0 : \rho = 0.9$

$H_1 : \rho \neq 0.9$

$$z_0 = (\text{arctan h} \quad R - \text{arctan h} \quad \rho_0)(n-3)^{1/2}$$

$$z_0 = (\text{arctan h} \quad 0.9975 - \text{arctan h} \quad 0.9)(12-3)^{1/2}$$

$$z_0 = 5.6084$$

$$z_{\alpha/2} = z_{0.025} = 1.96$$

$$|z_0| > z_{0.025}$$

Reject H_0, P-value ≈ 0

d)

$$\tanh\left(\text{arctan h} \quad r - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \le \rho \le \tanh\left(\text{arctan h} \quad r + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right)$$

$$\tanh\left(\text{arctan h} \quad 0.9975 - \frac{1.96}{\sqrt{12-3}}\right) \le \rho \le \tanh\left(\text{arctan h} \quad 0.9975 + \frac{19.6}{\sqrt{12-3}}\right)$$

$$0.9908 \le \rho \le 0.9993$$

11-85 Refer to the computer output in the referenced exercise.

a) r = $\sqrt{0.672}$ = 0.820

b) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.82\sqrt{32-2}}{\sqrt{1-0.82^2}} = 7.847$$

$$t_{0.025,30} = 2.042$$

$$t_0 > t_{0.025,30}$$

Reject H_0, P-value < 0.0005

c)

$$\tanh\left(\operatorname{arctan\,h}(0.082) - \frac{1.96}{\sqrt{32-3}}\right) \le \rho \le \tanh\left(\operatorname{arctan\,h}(0.082) + \frac{19.6}{\sqrt{32-3}}\right)$$

$$0.660 \le \rho \le 0.909$$

d)

$$H_0 : \rho = 0.7$$

$$H_1 : \rho \neq 0.7$$

$$z_0 = (\operatorname{arctan\,h} \quad R - \operatorname{arctan\,h} \quad \rho_0)(n-3)^{1/2}$$

$$z_0 = (\operatorname{arctan\,h} \quad 0.82 - \operatorname{arctan\,h} \quad 0.7)(32-3)^{1/2}$$

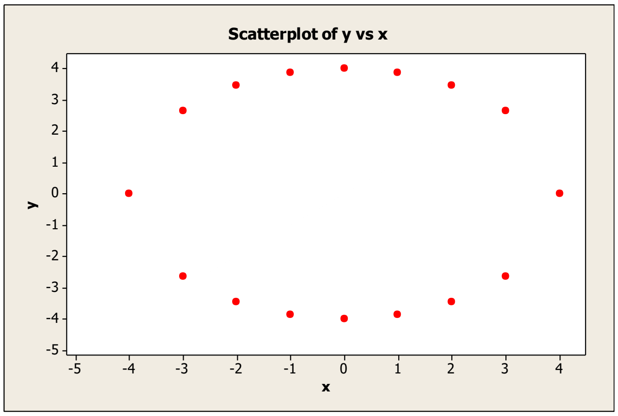
$$z_0 = 1.56$$

$$z_{\alpha/2} = z_{0.025} = 1.96$$

$$|z_0| < z_{0.025}$$

Fail to Reject H_0, P-value = 2(0.0594) = 0.119

11-86



Here r = 0. The correlation coefficient does not detect the relationship between x and y because the relationship is not linear. See the graph above.

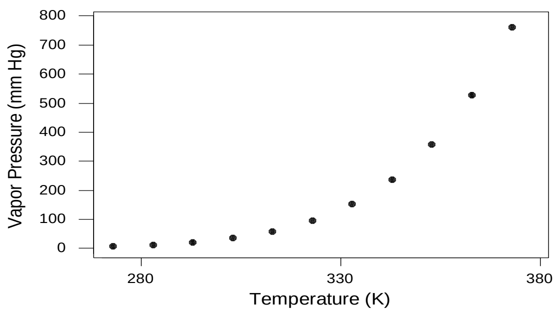
Section 11-9

11-87 a) Yes, $\ln y = \ln \beta_0 + \beta_1 \ln x + \ln \varepsilon$

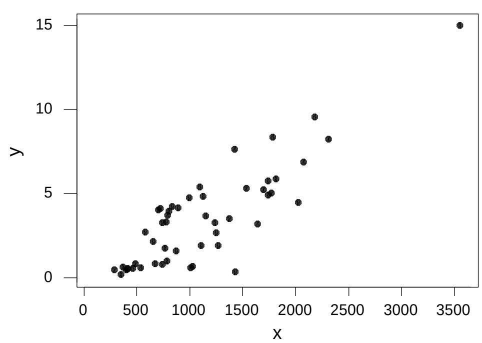
b) No

c) Yes, $\ln y = \ln \beta_0 + x \ln \beta_1 + \ln \varepsilon$

d) Yes, $\frac{1}{y} = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$



11-88 a) There is curvature in the data.

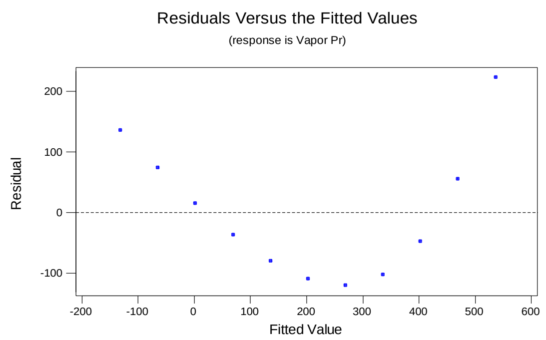


b) $y = -1956.3 + 6.686\,x$

c)

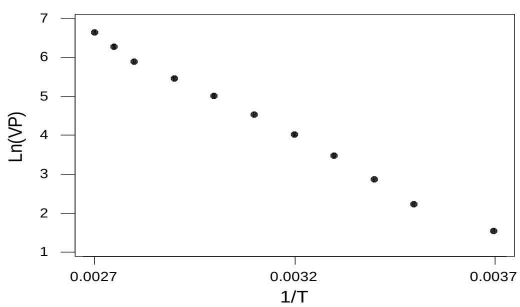
Source	DF	SS	MS	F	P
Regression	1	491662	491662	35.57	0.000
Residual Error	9	124403	13823		
Total	10	616065			

d)

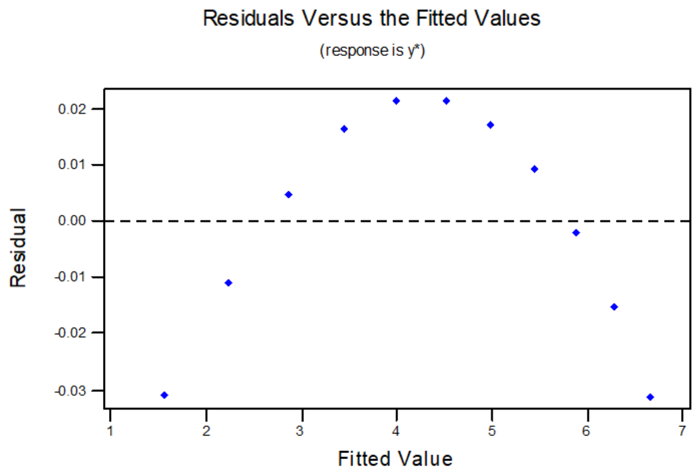


There is curvature in the plot of the residuals.

e) The data are linear after the transformation to $y^* = \ln y$ and $x^* = 1/x$.

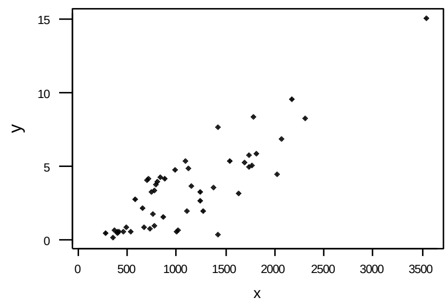



```
ln y = 20.6 - 5201(1/x)

Analysis of Variance
Source              DF          SS          MS          F        P
Regression           1      28.511      28.511   66715.47    0.000
Residual Error       9       0.004       0.000
Total               10      28.515
```



There is still curvature in the data, but now the plot is convex instead of concave.

11-89 a)



b) $\hat{y} = -0.8819 + 0.00385x$

c) $H_0 : \beta_1 = 0$

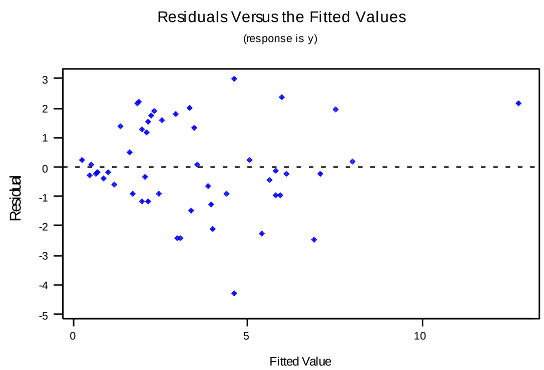
$H_1 : \beta_1 \neq 0$ $\alpha = 0.05$

$f_0 = 122.03$

$f_0 > f_{0.05,1,48}$

Reject H_0. Conclude that regression model is significant at $\alpha = 0.05$

d) No, it seems the variance is not constant, there is a funnel shape.



e) $\hat{y}^* = 0.5967 + 0.00097x$. Yes, the transformation stabilizes the variance.

Section 11-10

11-90 a) The fitted logistic regression model is $\hat{y} = \dfrac{1}{1+\exp[-(-8.73951-0.00020x)]}$

The computer results are shown below.

```
Binary Logistic Regression: Home Ownership Status versus Income

Link Function: Logit
Response Information
Variable               Value  Count
Home Ownership Status  1         11  (Event)
                       0          9
                       Total     20

Logistic Regression Table
                                                     Odds     95% CI
Predictor       Coef    SE Coef      Z      P  Ratio  Lower  Upper
Constant    -8.73951    4.43923  -1.97  0.049
Income     0.0002009  0.0001006   2.00  0.046   1.00   1.00   1.00

Log-Likelihood = -11.217
Test that all slopes are zero: G = 5.091, DF = 1, P-Value = 0.024
```

b) The P-value for the test of the coefficient of *income* is 0.046 < α = 0.05. Therefore, *income* has a significant effect on home ownership status.

c) The odds ratio is changed by the factor $\exp(\beta_1) = \exp(0.0002009) = 1.0002$ for every unit increase in *income*. More realistically, if income changes by \$1000, the odds ratio is changed by the factor $\exp(1000\beta_1) = \exp(0.2009) = 1.22$.

11-91 a) The fitted logistic regression model is $\hat{y} = \dfrac{1}{1+\exp[-(5.33971-0.00155x)]}$

The computer results are shown below.

```
Binary Logistic Regression: Number Failing, Sample Size, versus
Load, x(psi)

Link Function: Logit
Response Information
Variable           Value    Count
                   Success    353
Number Failing, r  Failure    337
Sample Size, n     Total      690

Logistic Regression Table
                                                         Odds     95% CI
Predictor           Coef    SE Coef      Z      P  Ratio  Lower  Upper
Constant         5.33971   0.545693   9.79  0.000
Load, x(psi)  -0.0015484  0.0001575  -9.83  0.000   1.00   1.00   1.00

Log-Likelihood = -421.856
Test that all slopes are zero: G = 112.460, DF = 1, P-Value = 0.000
```

b) The P-value for the test of the coefficient of *load* is near zero. Therefore, *load* has a significant effect on failing performance.

11-92 a) The fitted logistic regression model is $\hat{y} = \dfrac{1}{1+\exp[-(-2.08475+0.13573x)]}$

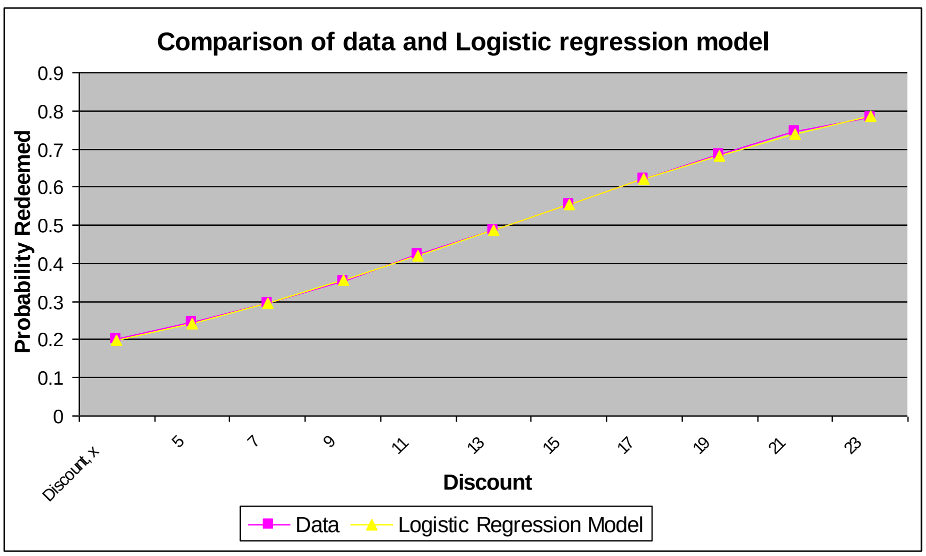
The computer results are shown below.

Binary Logistic Regression: Number Redee, Sample size, versus Discount, x

```
Link Function: Logit
Response Information
Variable             Value    Count
Number Redeemed, r   Success   2693
                     Failure   2807
Sample size,n        Total     5500

Logistic Regression Table
                                                      Odds      95% CI
Predictor         Coef     SE Coef       Z       P   Ratio  Lower  Upper
Constant      -2.08475   0.0803976  -25.93   0.000
Discount, x   0.135727   0.0049571   27.38   0.000    1.15   1.13   1.16

Log-Likelihood = -3375.665
Test that all slopes are zero: G = 870.925, DF = 1, P-Value = 0.000
```

b) The P-value for the test of the coefficient of *discount* is near zero. Therefore, *discount* has a significant effect on redemption.

c)



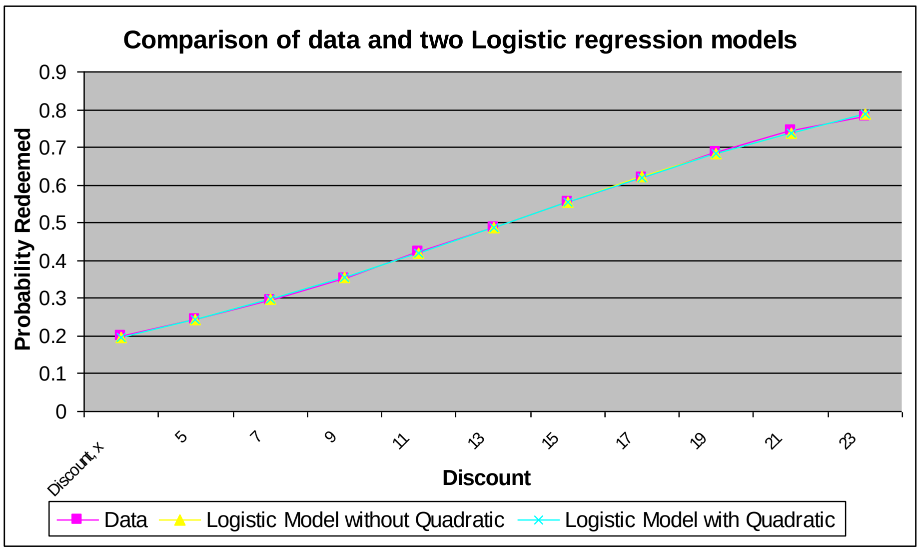
d) The P-value of the quadratic term is 0.95 > 0.05, so we fail to reject the null hypothesis of the quadratic coefficient at the 0.05 level of significance. There is no evidence that the quadratic term is required in the model. The computer results are shown below.

Binary Logistic Regression: Number Redee, Sample size, versus Discount, x

```
Link Function: Logit
Response Information
Variable             Value    Count
```

```
Number Redeemed, r  Success    2693
                    Failure    2807
Sample size,n       Total      5500

Logistic Regression Table
                                                                    95%
                                                             Odds    CI
Predictor                    Coef    SE Coef       Z      P  Ratio  Lower
Constant                 -2.07422   0.185045  -11.21  0.000
Discount, x              0.134072  0.0266622    5.03  0.000   1.14   1.09
Discount, x*Discount, x 0.0000550  0.0008707    0.06  0.950   1.00   1.00

Predictor               Upper
Constant
Discount, x              1.20
Discount, x*Discount, x  1.00

Log-Likelihood = -3375.663
Test that all slopes are zero: G = 870.929, DF = 2, P-Value = 0.000
```

e) The expanded model does not visually provide a better fit to the data than the original model.



11-93 a) The computer results are shown below.

Binary Logistic Regression: y versus Income x1, Age x2

```
Link Function: Logit
Response Information
Variable  Value   Count
y         1          10   (Event)
          0          10
          Total      20

Logistic Regression Table
                                                 Odds     95% CI
Predictor       Coef    SE Coef      Z      P  Ratio  Lower  Upper
Constant    -7.04706    4.67416  -1.51  0.132
Income x1  0.0000738  0.0000637   1.16  0.247   1.00   1.00   1.00
Age x2      0.987886   0.527358   1.87  0.061   2.69   0.96   7.55

Log-Likelihood = -10.541
```

```
Test that all slopes are zero: G = 6.644, DF = 2, P-Value = 0.036
```

b) Because the P-value = 0.036 < α = 0.05 we can conclude that at least one of the coefficients (of *income* and *age*) is not equal to zero at the 0.05 level of significance. The individual z-tests do not generate P-values less than 0.05, but this might be due to correlation between the independent variables. The z-test for a coefficient assumes it is the last variable to enter the model. A model might use either *income* or *age*, but after one variable is in the model, the coefficient z-test for the other variable may not significant because of their correlation.

c) The odds ratio is changed by the factor $\exp(\beta_1) = \exp(0.0000738) = 1.00007$ for every unit increase in *income* with *age* held constant. Similarly, odds ratio is changed by the factor $\exp(\beta_1) = \exp(0.987886) = 2.686$ for every unit increase in *age* with *income* held constant. More realistically, if income changes by $1000, the odds ratio is changed by the factor $\exp(1000\beta_1) = \exp(0.0738) = 1.077$ with *age* held constant.

d) At $x_1 = 45000$ and $x_2 = 5$ from part (a)

$$\hat{y} = \frac{1}{1+\exp[-(-7.04706+0.0000738x_1+0.987886x_2)]} = 0.77$$

e) The results from computer software are shown below.

```
Binary Logistic Regression: y versus Income x1, Age x2

Link Function: Logit
Response Information
Variable  Value  Count
y         1         10  (Event)
          0         10
          Total     20

Logistic Regression Table
                                                           Odds     95% CI
Predictor               Coef    SE Coef       Z       P  Ratio  Lower  Upper
Constant            0.314351    6.39401    0.05   0.961
Income x1         -0.0001411  0.0001412   -1.00   0.318   1.00   1.00   1.00
Age x2              -2.46169    2.08154   -1.18   0.237   0.09   0.00   5.04
Income x1*Age x2   0.0001014  0.0000630    1.61   0.107   1.00   1.00   1.00

Log-Likelihood = -8.275
Test that all slopes are zero: G = 11.175, DF = 3, P-Value = 0.011
```

Because the P-value = 0.107 there is no evidence that an interaction term is required in the model.

11-94

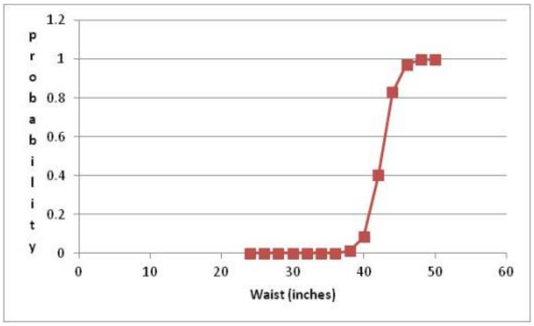
a) The logistic function is $= \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$. From a plot or from the derivative, this is a monotonic increasing function of x. Therefore, the probability increases as a function of x.

b) $p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{e^{-41.828+0.9864(36)}}{1+e^{-41.828+0.9864(36)}} = 0.0018$

c) $p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{e^{-41.828+0.9864(42)}}{1+e^{-41.828+0.9864(42)}} = 0.4015$

d) $p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{e^{-41.828+0.9864(48)}}{1+e^{-41.828+0.9864(48)}} = 0.9960$

e)



11-95

With failures defined at 2100 psi, the data is linearly separable with age and the logistic estimates do not converge. The exercise is changed so that data below 2300 psi is considered a failure. The computer output follows.

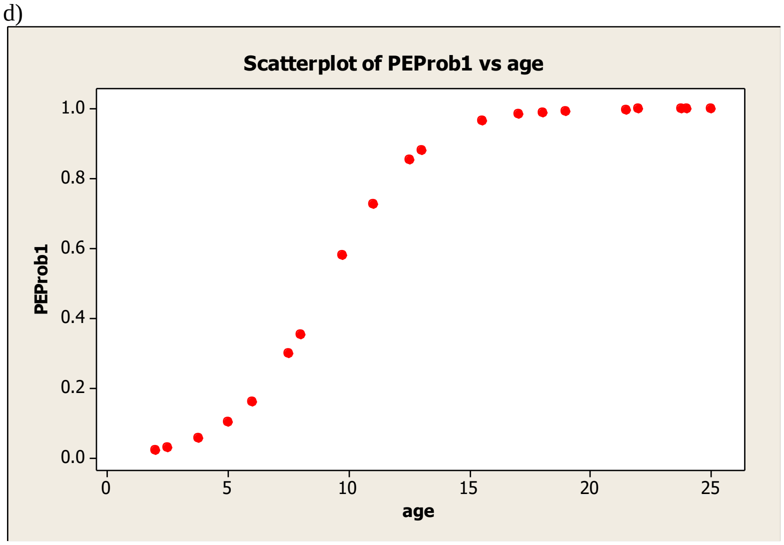
```
Response Information

Variable  Value  Count
failure   1         13  (Event)
          0          7
          Total     20

Logistic Regression Table

                                                    Odds     95% CI
Predictor       Coef   SE Coef      Z       P  Ratio  Lower  Upper
Constant   -4.78112   2.40936  -1.98  0.047
age        0.523235  0.251573   2.08  0.038   1.69   1.03   2.76

Log-Likelihood = -4.896
Test that all slopes are zero: G = 16.106, DF = 1, P-Value = 0.000
Goodness-of-Fit Tests

Method           Chi-Square  DF      P
Pearson             12.1296  18  0.840
Deviance             9.7920  18  0.938
Hosmer-Lemeshow      7.9635   8  0.437

Table of Observed and Expected Frequencies:
(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

                             Group
Value    1    2    3    4    5    6    7    8    9   10  Total
1
  Obs    0    1    0    0    2    2    2    2    2    2     13
  Exp  0.1  0.2  0.5  0.9  1.6  1.8  2.0  2.0  2.0  2.0
0
  Obs    2    1    2    2    0    0    0    0    0    0      7
  Exp  1.9  1.8  1.5  1.1  0.4  0.2  0.0  0.0  0.0  0.0
Total    2    2    2    2    2    2    2    2    2    2     20

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs        Number  Percent  Summary Measures
Concordant       87     95.6  Somers' D              0.91
Discordant        4      4.4  Goodman-Kruskal Gamma  0.91
Ties              0      0.0  Kendall's Tau-a        0.44
Total            91    100.0
```

a) The P-value for the test that the coefficient of age is zero ($\beta_1 = 0$) is 0.038. Therefore, age has a significant effect on the probability of failure at $\alpha = 0.05$.

b) The fitted model is log[p/(1-p)] = -4.78 + 0.523x. At x= 18, p/(1-p) = exp[-4.78 + 0.523(18)] and p = 0.990.

c) The odds of failure are p/(1-p) = exp[-4.78 + 0.523x]. A one week increase in age changes the odds to
p/(1-p) = exp[-4.78 + 0.523(x+1)] = exp[-4.78 + 0.523x]exp(0.523)
Therefore, the odds are multiplied by exp(0.523) = 1.69.

d)



Supplemental Exercises

11-96 a) $\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{y}_i$ and $\sum y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i$ from the normal equations

Then,

$$(n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i) - \sum \hat{y}_i$$

$$= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0$$

b) $\sum_{i=1}^{n}(y_i - \hat{y}_i)x_i = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n}\hat{y}_i x_i$

and $\sum_{i=1}^{n} y_i x_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2$ from the normal equations. Then,

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i)x_i =$$

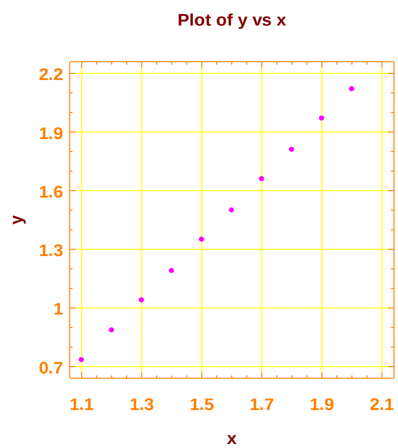
$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

c) $\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \bar{y}$

$$\sum \hat{y} = \sum(\hat{\beta}_0 + \hat{\beta}_1 x)$$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i &= \frac{1}{n}\sum(\hat{\beta}_0 + \hat{\beta}_1 x_i) \\
&= \frac{1}{n}(n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i) \\
&= \frac{1}{n}(n(\bar{y} - \hat{\beta}_1\bar{x}) + \hat{\beta}_1 \sum x_i) \\
&= \frac{1}{n}(n\bar{y} - n\hat{\beta}_1\bar{x} + \hat{\beta}_1 \sum x_i) \\
&= \bar{y} - \hat{\beta}\bar{x} + \hat{\beta}_1\bar{x} \\
&= \bar{y}
\end{aligned}$$

11-97 a)



Yes, a linear relationship seems plausible.

b)

```
                         Model fitting results for: y
Independent variable                coefficient  std. error     t-value   sig.level
CONSTANT                              -0.966824    0.004845   -199.5413      0.0000
x                                      1.543758    0.003074    502.2588      0.0000
--------------------------------------------------------------------------------
R-SQ. (ADJ.) = 1.0000   SE=        0.002792  MAE=        0.002063  DurbWat=  2.843
Previously:    0.0000              0.000000              0.000000            0.000
10 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

$\hat{y} = -0.966824 + 1.54376x$

c)

```
                 Analysis of Variance for the Full Regression
Source                 Sum of Squares      DF     Mean Square       F-Ratio   P-value
Model                         1.96613       1         1.96613       252264.     .0000
Error                   0.0000623515        8  0.00000779394
--------------------------------------------------------------------------------
Total (Corr.)                 1.96619       9
R-squared = 0.999968                              Stnd. error of est. = 2.79176E-3
R-squared (Adj. for d.f.) = 0.999964           Durbin-Watson statistic = 2.84309
```

2) $H_0: \beta_1 = 0$

3) $H_1: \beta_1 \neq 0$

4) $\alpha = 0.05$

5) The test statistic is $f_0 = \dfrac{SS_R / k}{SS_E /(n-p)}$

6) Reject H_0 if $f_0 > f_{\alpha,1,8}$ where $f_{0.05,1,8} = 5.32$
7) Using the results from the ANOVA table

$$f_0 = \frac{1.96613/1}{0.0000623515/8} = 252263.9$$

8) Because 252264 > 5.32 reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$.
P-value ≈ 0

d)

```
          95 percent confidence intervals for coefficient estimates
  --------------------------------------------------------------------------------
                         Estimate  Standard error     Lower Limit     Upper Limit
  CONSTANT                -0.96682        0.00485        -0.97800        -0.95565
  x                        1.54376        0.00307         1.53667         1.55085
  --------------------------------------------------------------------------------
```

$$1.53667 \le \beta_1 \le 1.55085$$

e) 2) $H_0: \beta_0 = 0$
3) $H_1: \beta_0 \neq 0$
4) $\alpha = 0.05$
5) The test statistic is $t_0 = \dfrac{\hat{\beta}_0}{se(\hat{\beta}_0)}$
6) Reject H_0 if $t_0 < -t_{\alpha/2,n-2}$ where $-t_{0.025,8} = -2.306$ or $t_0 > t_{0.025,8} = 2.306$
7) Using the results from the table above

$$t_0 = \frac{-0.96682}{0.00485} = -199.34$$

8) Because −199.34 < −2.306 reject H_0 and conclude the intercept is significant at $\alpha = 0.05$.

11-98 a) $\hat{y} = 93.34 + 15.64x$

b) $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$ $\alpha = 0.05$

$f_0 = 12.872$

$f_{0.05,1,14} = 4.60$

$f_0 > f_{0.05,1,14}$

Reject H_0. Conclude that $\beta_1 \neq 0$ at $\alpha = 0.05$.

c) $(7.961 \le \beta_1 \le 23.322)$

d) $(74.758 \le \beta_0 \le 111.923)$

e) $\hat{y} = 93.34 + 15.64(2.5) = 132.44$

$$132.44 \pm 2.145\sqrt{136.27\left[\tfrac{1}{16} + \tfrac{(2.5-2.325)^2}{7.017}\right]}$$

$$132.44 \pm 6.468$$

$$125.97 \le \hat{\mu}_{Y|x_0=2.5} \le 138.91$$

11-99 $\hat{y}^* = 1.2232 + 0.5075x$ where $y^* = 1/y$. No, the model does not seem reasonable.
The residual plots indicate a possible outlier.

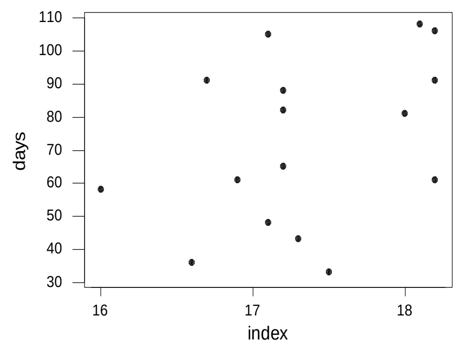
11-100 $\hat{y} = 4.5755 + 2.2047x$, $r = 0.992$, $R^2 = 98.40\%$

The model appears to be a good fit. The R^2 is large and both regression coefficients are significant. No, the existence of a strong correlation does not imply a cause and effect relationship.

11-101 $\hat{y} = 0.7916x$

Even though y should be zero when x is zero, because the regressor variable does not usually assume values near zero, a model with an intercept fits this data better. Without an intercept, the residuals plots are not satisfactory.

11-102 a)



b) The regression equation is

$$\hat{y} = -193 + 15.296x$$

Analysis of Variance

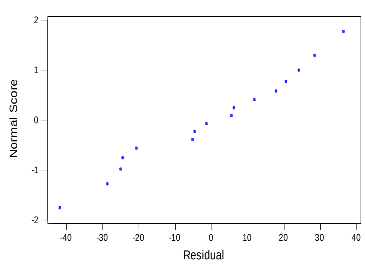
Source	DF	SS	MS	F	P
Regression	1	1492.6	1492.6	2.64	0.127
Residual Error	14	7926.8	566.2		
Total	15	9419.4			

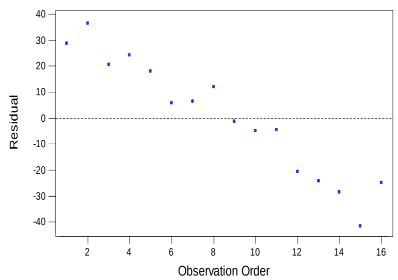
Fail to reject H_0. We do not have evidence of a relationship. Therefore, there is not sufficient evidence to conclude that the seasonal meteorological index (*x*) is a reliable predictor of the number of days that the ozone level exceeds 0.20 ppm (*y*).

c) 95% CI on β_1

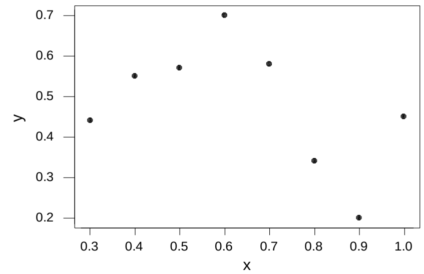
$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} se(\hat{\beta}_1)$$
$$15.296 \pm t_{.025,12}(9.421)$$
$$15.296 \pm 2.145(9.421)$$
$$-4.912 \le \beta_1 \le 35.504$$

d) The normality plot of the residuals is satisfactory. However, the plot of residuals versus run order exhibits a strong downward trend. This could indicate that there is another variable should be included in the model and it is one that changes with time.





11-103 a)



b) $\hat{y} = 0.6714 - 2964x$

c)

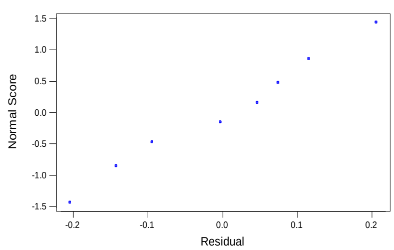
Analysis of Variance

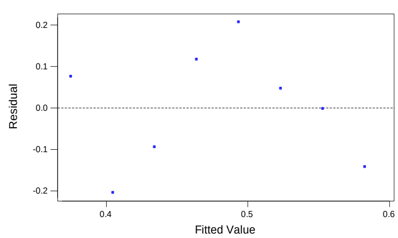
Source	DF	SS	MS	F	P
Regression	1	0.03691	0.03691	1.64	0.248
Residual Error	6	0.13498	0.02250		
Total	7	0.17189			

$R^2 = 21.47\%$

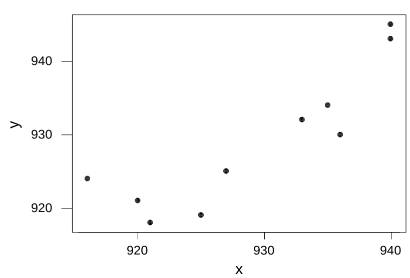
Because the P-value > 0.05, reject the null hypothesis and conclude that the model is significant.

d) There appears to be curvature in the data. There is a dip in the middle of the normal probability plot and the plot of the residuals versus the fitted values shows curvature.





11-104 a)



b) $\hat{y} = 33.3 + 0.9636x$

c)

```
Predictor        Coef     SE Coef        T        P
Constant         66.0       194.2     0.34    0.743
Therm          0.9299      0.2090     4.45    0.002

S = 5.435       R-Sq = 71.2%     R-Sq(adj) = 67.6%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      584.62      584.62     19.79    0.002
Residual Error     8      236.28       29.53
Total              9      820.90
```

Reject the null hypothesis and conclude that the model is significant. Here 77.3% of the variability is explained by the model.

d) $H_0 : \beta_1 = 1$

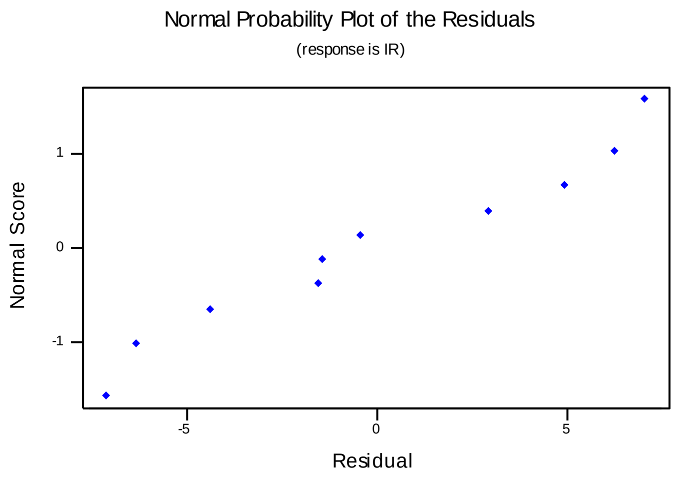
$H_1 : \beta_1 \neq 1$ $\alpha = 0.05$

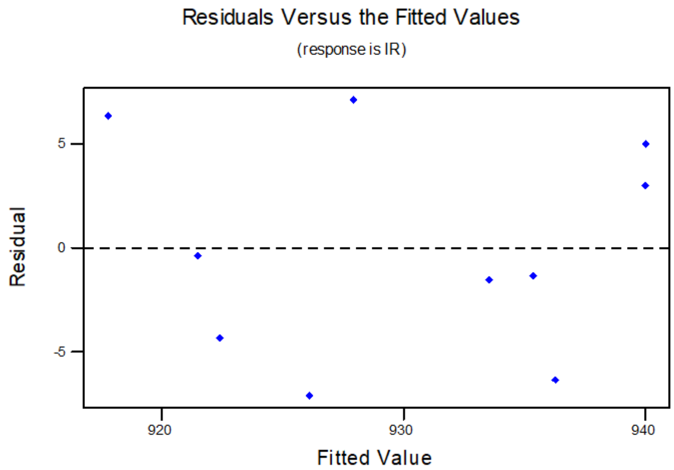
$$t_0 = \frac{\hat{\beta}_1 - 1}{se(\hat{\beta}_1)} = \frac{0.9299 - 1}{0.2090} = -0.3354$$

$$t_{a/2,n-2} = t_{.025,8} = 2.306$$

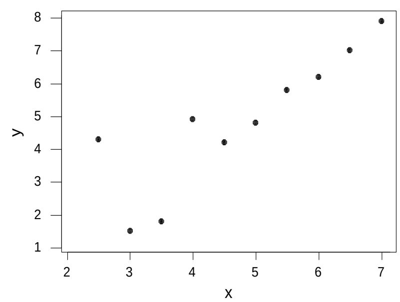
Because $t_0 > -t_{a/2,n-2}$, we fail to reject H_0. There is not enough evidence to reject the claim that the devices produce different temperature measurements.

e) The residual plots to not reveal any major problems.





11-105 a)



b) $\hat{y} = -0.699 + 1.66x$

c)

Source	DF	SS	MS	F	P
Regression	1	28.044	28.044	22.75	0.001
Residual Error	8	9.860	1.233		
Total	9	37.904			

Reject the null hypothesis and conclude that the model is significant.

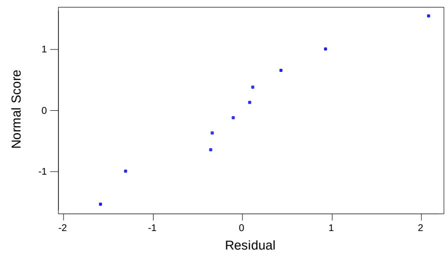
d) x_0 = 4.25 $\hat{\mu}_{y|x_0} = 4.257$

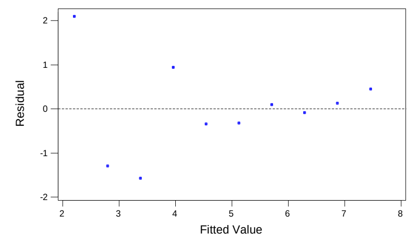
$$4.257 \pm 2.306\sqrt{1.2324\left(\frac{1}{10} + \frac{(4.25-4.75)^2}{20.625}\right)}$$

$$4.257 \pm 2.306(0.3717)$$

$$3.399 \le \mu_{y|x_o} \le 5.114$$

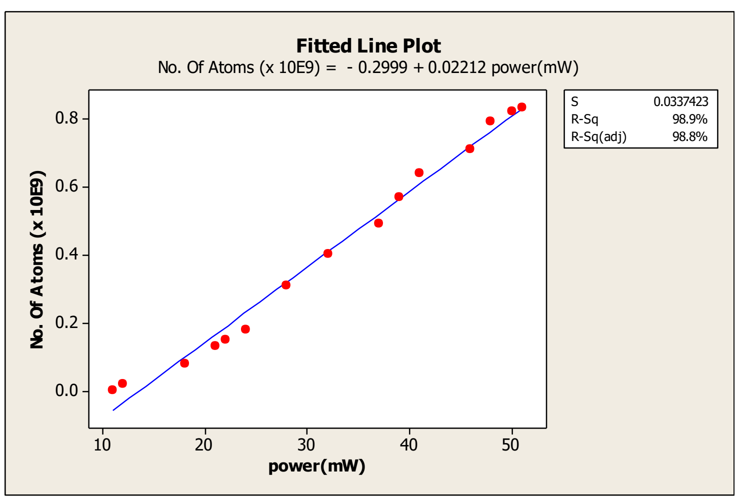
e) The normal probability plot of the residuals appears linear, but there are some large residuals in the lower fitted values. There may be some problems with the model.





11-106 a)

```
The regression equation is
No. Of Atoms (x 10E9) = - 0.300 + 0.0221 power(mW)


Predictor        Coef     SE Coef       T      P
Constant     -0.29989     0.02279  -13.16  0.000
power(mW)   0.0221217   0.0006580   33.62  0.000


S = 0.0337423   R-Sq = 98.9%   R-Sq(adj) = 98.8%


Analysis of Variance

Source          DF      SS      MS        F      P
Regression       1  1.2870  1.2870  1130.43  0.000
Residual Error  13  0.0148  0.0011
Total           14  1.3018
```



b) Yes, there is a significant regression at $\alpha = 0.05$ because p-value = 0.000 < α.

c) $r = \sqrt{0.989} = 0.994$

d)

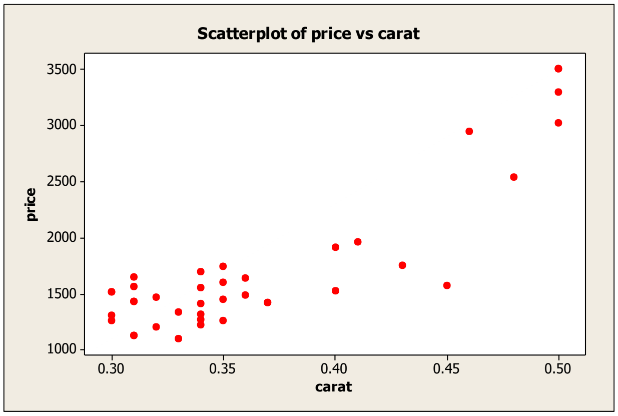
$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$
$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.994\sqrt{15-2}}{\sqrt{1-.994^2}} = 32.766$$
$$t_{0.025,13} = 2.160$$
$$t_0 = 32.766 > t_{0.025,13} = 2.160.$$

Reject H_0, P-value ≈ 0.000

e) 95% confidence interval for $\hat{\beta}_1$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2}se(\hat{\beta}_1)$$
$$0.022 \pm t_{0.025,13}(0.00066)$$
$$0.022 \pm 2.160(0.00066)$$
$$0.0206 \leq \hat{\beta}_1 \leq 0.0234$$

11-107 a)



The relationship between carat and price is not linear. Yes, there is one outlier, observation number 33.

b) The person obtained a very good price—high carat diamond at low price.

c) All the data

```
The regression equation is
price = - 1696 + 9349 carat

Predictor      Coef  SE Coef      T      P
Constant    -1696.2    298.3  -5.69  0.000
carat        9349.4    794.1  11.77  0.000

S = 331.921   R-Sq = 78.5%   R-Sq(adj) = 77.9%

Analysis of Variance

Source          DF        SS        MS       F      P
Regression       1  15270545  15270545  138.61  0.000
Residual Error  38   4186512    110171
Total           39  19457057
```

$t_{\alpha/2,n-2} = t_{0.025,38} = 2.024$
95% confidence interval on β_1

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2}se(\hat{\beta}_1)$$
$$9349 \pm t_{.025,38}(794.1)$$
$$9349 \pm 2.024(794.1)$$
$$7741.74 \le \beta_1 \le 10956.26.$$

With unusual data omitted

```
The regression equation is
price_1 = - 1841 + 9809 carat_1


Predictor     Coef  SE Coef      T      P
Constant   -1841.2    269.9  -6.82  0.000
carat_1     9809.2    722.5  13.58  0.000


S = 296.218   R-Sq = 83.3%   R-Sq(adj) = 82.8%


Analysis of Variance

Source          DF        SS        MS       F      P
Regression       1  16173949  16173949  184.33  0.000
Residual Error  37   3246568     87745
Total           38  19420517
```

$t_{\alpha/2,n-2} = t_{0.025,37} = 2.026$
95% confidence interval on β_1.

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2}se(\hat{\beta}_1)$$
$$9809 \pm t_{.025,37}(722.5)$$
$$9809 \pm 2.026(722.5)$$
$$8345.22 \le \beta_1 \le 11272.79$$

The width for the outlier removed is narrower than for the first case.

11-108

```
The regression equation is
Population = 3549143 + 651828 Count


Predictor     Coef  SE Coef      T      P
Constant   3549143   131986  26.89  0.000
Count       651828   262844   2.48  0.029


S = 183802   R-Sq = 33.9%   R-Sq(adj) = 28.4%


Analysis of Variance

Source          DF           SS           MS     F      P
Regression       1  2.07763E+11  2.07763E+11  6.15  0.029
Residual Error  12  4.05398E+11  33783126799
Total           13  6.13161E+11
```

$$\hat{y} = 3549143 + 651828x$$

Yes, the regression is significant at α = 0.05. Care needs to be taken in making cause and effect statements based on a regression analysis. In this case, it is surely not the case that an increase in the stork count is causing the population to increase, in fact, the opposite is most likely the case. However, unless a designed experiment is performed, cause and effect statements should not be made on regression analysis alone. The existence of a strong correlation does not imply a cause and effect relationship.

Mind-Expanding Exercises

11-109 The correlation coefficient for the *n* pairs of data (x_i, z_i) can be much different from unity. For example, if y = bx and if the x data is symmetric about zero, the correlation coefficient between x and y^2 is zero. In other cases, it can be much less than unity (in absolute value). Over some restricted ranges of x values, the quadratic function $y = (a + bx)^2$ can be approximated by a linear function of x and in these cases the correlation can still be near unity. However, in general, the correlation can be much different from unity and in some cases equal zero. Correlation is a measure of a linear relationship, and if a nonlinear relationship exists between variables, even if it is strong, the correlation coefficient does not usually provide a good measure.

11-110 a) $\hat{\beta}_1 = \dfrac{S_{xY}}{S_{xx}}$, $\quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \bar{x}$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\overline{Y}, \hat{\beta}_1) - \bar{x}Cov(\hat{\beta}_1, \hat{\beta}_1)$$

$$Cov(\overline{Y}, \hat{\beta}_1) = \frac{Cov(\overline{Y}, S_{xY})}{S_{xx}} = \frac{Cov(\sum Y_i, \sum Y_i(x_i - \bar{x}))}{nS_{xx}} = \frac{\sum (x_i - \bar{x})\sigma^2}{nS_{xx}} = 0.$$ Therefore,

$$Cov(\hat{\beta}_1, \hat{\beta}_1) = V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}\sigma^2}{S_{xx}}$$

b) The requested result is shown in part a).

11-111 a) $MS_E = \dfrac{\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2} = \dfrac{\sum e_i^2}{n-2}$

$$E(e_i) = E(Y_i) - E(\hat{\beta}_0) - E(\hat{\beta}_1)x_i = 0$$

$$V(e_i) = \sigma^2[1 - (\tfrac{1}{n} + \tfrac{(x_i - \bar{x})^2}{S_{xx}})]$$ Therefore,

$$E(MS_E) = \frac{\sum E(e_i^2)}{n-2} = \frac{\sum V(e_i)}{n-2}$$

$$= \frac{\sum \sigma^2[1 - (\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}})]}{n-2}$$

$$= \frac{\sigma^2[n-1-1]}{n-2} = \sigma^2$$

b) Using the fact that $SS_R = MS_R$, we obtain

$$E(MS_R) = E(\hat{\beta}_1^2 S_{xx}) = S_{xx}\left\{V(\hat{\beta}_1) + [E(\hat{\beta}_1)]^2\right\}$$

$$= S_{xx}\left(\frac{\sigma^2}{S_{xx}} + \beta_1^2\right) = \sigma^2 + \beta_1^2 S_{xx}$$

11-112 $\hat{\beta}_1 = \dfrac{S_{x_1Y}}{S_{x_1x_1}}$

$$E(\hat{\beta}_1) = \frac{E\left[\sum_{i=1}^{n} Y_i(x_{1i} - \bar{x}_1)\right]}{S_{x_1x_1}} = \frac{\sum_{i=1}^{n}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})(x_{1i} - \bar{x}_1)}{S_{x_1x_1}}$$

$$= \frac{\beta_1 S_{x_1x_1} + \beta_2 \sum_{i=1}^{n} x_{2i}(x_{1i} - \bar{x}_1)}{S_{x_1x_1}} = \beta_1 + \frac{\beta_2 S_{x_1x_2}}{S_{x_1x_1}}$$

No, $\hat{\beta}_1$ is no longer unbiased.

11-113 $V(\hat{\beta}_1) = \dfrac{\sigma^2}{S_{xx}}$. To minimize $V(\hat{\beta}_1)$, S_{xx} should be maximized.

Because $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$, S_{xx} is maximized by choosing approximately half of the observations at each end of the range of *x*. From a practical perspective, this allocation assumes the linear model between *Y* and *x* holds throughout the range of *x* and observing *Y* at only two *x* values prohibits verifying the linearity assumption. It is often preferable to obtain some observations at intermediate values of *x*.

11-114 One might minimize a weighted some of squares $\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2$ in which a Y_i with small variance (w_i large) receives greater weight in the sum of squares.

$$\frac{\partial}{\beta_0}\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2 = -2\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{\partial}{\beta_1}\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)^2 = -2\sum_{i=1}^{n} w_i(y_i - \beta_0 - \beta_1 x_i)x_i$$

Setting these derivatives to zero yields

$$\hat{\beta}_0 \sum w_i + \hat{\beta}_1 \sum w_i x_i = \sum w_i y_i$$

$$\hat{\beta}_0 \sum w_i x_i + \hat{\beta}_1 \sum w_i x_i^2 = \sum w_i x_i y_i$$

and these equations are solved as follows

$$\hat{\beta}_1 = \frac{\left(\sum w_i x_i y_i\right)\left(\sum w_i\right) - \sum w_i y_i}{\left(\sum w_i\right)\left(\sum w_i x_i^2\right) - \left(\sum w_i x_i\right)^2}$$

$$\hat{\beta}_0 = \frac{\sum w_i y_i}{\sum w_i} - \frac{\sum w_i x_i}{\sum w_i}\hat{\beta}_1 \quad .$$

11-115 $\hat{y} = \bar{y} + r\dfrac{s_y}{s_x}(x - \bar{x})$

$$= \bar{y} + \frac{S_{xy}\sqrt{\sum(y_i - \bar{y})^2}\,(x - \bar{x})}{\sqrt{S_{xx}S_{yy}}\sqrt{\sum(x_i - \bar{x})^2}}$$

$$= \bar{y} + \frac{S_{xy}}{S_{xx}}(x - \bar{x})$$

$$= \bar{y} + \hat{\beta}_1 x - \hat{\beta}_1\bar{x} = \hat{\beta}_0 + \hat{\beta}_1 x$$

11-116 a) $\dfrac{\partial}{\beta_1}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i$

Upon setting the derivative to zero, we obtain

$$\beta_0\sum x_i + \beta_1\sum x_i^{\,2} = \sum x_i y_i$$

Therefore,

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \beta_0\sum x_i}{\sum x_i^{\,2}} = \frac{\sum x_i(y_i - \beta_0)}{\sum x_i^{\,2}}$$

b) $V(\hat{\beta}_1) = V\left(\dfrac{\sum x_i\ (Y_i - \beta_0)}{\sum x_i^{\,2}}\right) = \dfrac{\sum x_i^{\,2}\sigma^2}{[\sum x_i^{\,2}]^2} = \dfrac{\sigma^2}{\sum x_i^{\,2}}$

c) $\hat{\beta}_1 \pm t_{\alpha/2,n-1}\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^{\,2}}}$

This confidence interval is shorter because $\sum x_i^{\,2} \geq \sum (x_i - \bar{x})^2$. Also, the *t* value based on *n* – 1 degrees of freedom is slightly smaller than the corresponding *t* value based on *n* – 2 degrees of freedom.

CHAPTER 12

Sections 12-1

12-1 Exercise 11.1 described a regression model between percent of body fat (%BF) as measured by immersion and BMI from a study on 250 male subjects. The researchers also measured 13 physical characteristics of each man, including his age (yrs), height (in), and waist size (in).

A regression of percent of body fat with both height and waist as predictors shows the following computer output:

	Estimate	Std. Error	t-value	Pr (>|t|)
(Intercept)	-3.10088	7.68611	-0.403	0.687
Height	-0.60154	0.10994	-5.472	1.09e-07
Waist	1.77309	0.07158	24.770	< 2e-16

Residual standard error: 4.46 on 247 degrees of freedom
Multiple R-squared: 0.7132, Adjusted R-squared: 0.7109
F-statistic: 307.1 on 2 and 247 DF, p-value: < 2.2e-16

(a) Write out the regression model if

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 2.9705 & -4.0042E\text{-}2 & -4.1679E\text{-}2 \\ -0.04004 & 6.0774E\text{-}4 & -7.3875E\text{-}5 \\ -0.00417 & -7.3875E\text{-}5 & 2.5766E\text{-}4 \end{bmatrix}$$

and

$$(\mathbf{X}'\mathbf{y}) = \begin{bmatrix} 4757.9 \\ 334335.8 \\ 179706.7 \end{bmatrix}$$

(b) Verify that the model found from technology is correct to at least 2 decimal places.

(c) What is the predicted body fat of a man who is 6-ft tall with a 34-in waist?

The entry in row 1, column 3 and row 3, column 1 of $(X'X)^{-1}$ should be -4.1679E-3.
(a)
$(X'X)^{-1}$ =

2.9705	-0.04004	-0.00417
-0.04004	0.000608	-7.4E-05
-0.00417	-7.4E-05	0.000258

X'y =

4757.9
334335.8
179706.7

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} -3.13171 \\ -0.59291 \\ 1.77372 \end{pmatrix}$$

$\hat{y} = -3.10088 - 0.60154x_1 + 1.77309x_2$ where x_1 = height and x_2 = weight

(b) $\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} -3.13171 \\ -0.59291 \\ 1.77372 \end{pmatrix} \cong \begin{pmatrix} -3.10 \\ -0.60 \\ -1.77 \end{pmatrix}$

This result agrees with the previous model up to two decimal places.

(c)

$$\hat{y} = -3.10088 - 0.60154x_1 + 1.77309x_2$$
$$\hat{y} = -3.10088 - 0.60154(72) + 1.77309(34)$$
$$\hat{y} = 13.87$$

12-2 A class of 63 students has two hourly exams and a final exam. How well do the two hourly exams predict performance on the final?

The following are some quantities of interest:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.9129168 & -9.815022e{-}03 & -7.11238e{-}04 \\ -0.00981502 & 1.497241e{-}04 & -4.15806e{-}05 \\ -0.00071123 & -4.15805e{-}05 & 5.81235e{-}05 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{y}) = \begin{bmatrix} 4871.0 \\ 426011.0 \\ 367576.5 \end{bmatrix}$$

(a) Calculate the least squares estimates of the slopes for hourly 1 and hourly 2 and the intercept.
(b) Use the equation of the fitted line to predict the final exam score for a student who scored 70 on hourly 1 and 85 on hourly 2.
(c) If a student who scores 80 on hourly 1 and 90 on hourly 2 gets an 85 on the final, what is her residual?

$(X'X)^{-1}$ =

0.912917	-9.82E-03	-7.11E-04
-0.00982	1.50E-04	-4.16E-05
-7.11E-04	-4.16E-05	5.81E-05

X'y =

4871
426011
367576.5

(a) $\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} 4.076021 \\ 0.691136 \\ 0.186642 \end{pmatrix}$

(b) $\hat{y} = 4.076021 + 0.691136(70) \quad + 0.186642(85) = 68.32$

(c) $\hat{y} = 4.076021 + 0.691136(80) \quad + 0.186642(90) = 76.16$

$$e = y - \hat{y} = 85 - 76.16 = 8.84$$

12-3 Can the percentage of the workforce who are engineers in each U.S. state be predicted by the amount of money spent in on higher education (as a percent of gross domestic product), on venture capital (dollars per $1000 of gross domestic product) for high-tech business ideas, and state funding (in dollars per student) for major research universities? Data for all 50 states and a software package revealed the following results:

Estimate	Std. Error	t value	Pr (>|t|)

(Intercept)	1.051e+00	1.567e-01	6.708	2.5e-08 ***
Venture cap	9.514e-02	3.910e-02	2.433	0.0189 *
State funding	4.106e-06	1.437e-05	0.286	0.7763
Higher.eD	-1.673e-01	2.595e-01	-0.645	0.5223

Residual standard error: 0.3007 on 46 degrees of freedom
Multiple R-squared: 0.1622, Adjusted R-squared: 0.1075
F-statistic: 2.968 on 3 and 46 DF, p-value: 0.04157

(a) Write the equation predicting the percent of engineers in the workforce.
(b) For a state that has $1 per $1000 in venture capital, spends $10,000 per student on funding for major research universities, and spends 0.5% of its GDP on higher education, what percent of engineers do you expect to see in the workforce?
(c) If the state in part (b) actually had 1.5% engineers in the workforce, what would the residual be?

(a) $\hat{y} = 1.051 + 0.09514x_1 + 0.000004106x_2 - 0.1673x_3$ where

x_1 : Venture capital, x_2 : State funding, x_3 : Higher education

(b) $\hat{y} = 1.051 + 0.09514x_1 + 0.000004106x_2 - 0.1673x_3$

$$\hat{y} = 1.051 + 0.09514 \times 1 + 0.000004106 \times 10000 - 0.1673 \times 0.5$$
$$= 1.08302$$

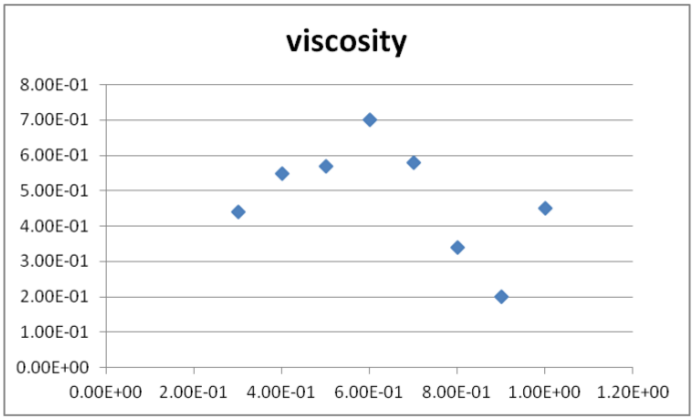
(c) $e = y - \hat{y} = 1.5 - 1.08302 = 0.41698$

12-4 Hsuie, Ma, and Tsai (‘‘Separation and Characterizations of Thermotropic Copolyesters of p-Hydroxybenzoic Acid, Sebacic Acid, and Hydroquinone,’’ (1995, Vol. 56) studied the effect of the molar ratio of sebacic acid (the regressor) on the intrinsic viscosity of copolyesters (the response). The following display presents the data.

Ratio	Viscosity
1.0	0.45
0.9	0.20
0.8	0.34
0.7	0.58
0.6	0.70
0.5	0.57
0.4	0.55
0.3	0.44

(a) Construct a scatterplot of the data.
(b) Fit a second-order prediction equation.

(a)



(b) To fit a second order model, the ratio term is squared

intercept	ratio	ratio2	viscosity
1.00E+00	1.00E+00	1.00E+00	4.50E-01
1	0.9	8.10E-01	0.2
1	0.8	6.40E-01	0.34
1	0.7	4.90E-01	0.58
1	0.6	3.60E-01	0.7
1	0.5	2.50E-01	0.57
1	0.4	1.60E-01	0.55
1	0.3	9.00E-02	0.44

$(X'X)^{-1} =$

1.872356	-6.67954	8.055106
-6.67954	36.28527	-50.277
8.055106	-50.277	73.45787

$X'y =$

3.83
2.365
1.6359

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} 0.197917 \\ 1.367262 \\ -1.27976 \end{pmatrix}$$

$$\hat{y} = 0.197917 + 1.367262x \; - 1.27976x^2$$

12-5 A study was performed to investigate the shear strength of soil (y) as it related to depth in feet (x_1) and percent of moisture content (x_2). Ten observations were collected, and the following summary quantities obtained: $n = 10$, $\sum x_{i1} = 223$, $\sum x_{i2} = 553$, $\sum y_i = 1{,}916$, $\sum x_{i1}^2 = 5{,}200.9$, $\sum x_{i2}^2 = 31{,}729$, $\sum x_{i1}x_{i2} = 12{,}352$, $\sum x_{i1}y_i = 43{,}550.8$, $\sum x_{i2}y_i = 104{,}736.8$, and $\sum y_i^2 = 371{,}595.6$.

(a) Set up the least squares normal equations for the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$.

(b) Estimate the parameters in the model in part (a).
(c) What is the predicted strength when x_1 =18 feet and x_2 = 43%?

(a) $X'X = \begin{bmatrix} 10 & 223 & 553 \\ 223 & 5200.9 & 12352 \\ 553 & 12352 & 31729 \end{bmatrix}$

$X'y = \begin{bmatrix} 1916.0 \\ 43550.8 \\ 104736.8 \end{bmatrix}$

(b) $\hat{\beta} = \begin{bmatrix} 171.05545 \\ 3.71329 \\ -1.12589 \end{bmatrix}$ so $\hat{y} = 171.055 + 3.713x_1 - 1.126x_2$

(c) $\hat{y} = 171.055 + 3.713(18) - 1.126(43) = 189.49$

12-6 A regression model is to be developed for predicting the ability of soil to absorb chemical contaminants. Ten observations have been taken on a soil absorption index (y) and two regressors: x_1 = amount of extractable iron ore and x_2 = amount of bauxite. We wish to fi t the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$. Some necessary quantities are:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1.17991 & -7.30982\text{ E-3} & -7.3006\text{ E-4} \\ -7.30982\text{ E-3} & 7.9799\text{ E-5} & -1.23713\text{ E-4}, \\ -7.3006\text{ E-4} & -1.23713\text{ E-4} & 4.6576\text{ E-4} \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{y}) = \begin{bmatrix} 220 \\ 36{,}768 \\ 9{,}965 \end{bmatrix}$$

(a) Estimate the regression coefficients in the model specified.
(b) What is the predicted value of the absorption index y when x_1 = 200 and x_2 = 50?

(a) $\hat{\beta} = (X'X)^{-1}X'y$

$$\hat{\beta} = \begin{bmatrix} -1.9122 \\ 0.0931 \\ 0.2532 \end{bmatrix}$$

(b) $\hat{y} = -1.9122 + 0.0931x_1 + 0.2532x_2$
$\hat{y} = -1.9122 + 0.0931(200) + 0.2532(50) = 29.3678$

12-7 A chemical engineer is investigating how the amount of conversion of a product from a raw material (y) depends on reaction temperature (x_1) and the reaction time (x_2). He has developed the following regression models:

1. $\hat{y} = 100 + 2x_1 + 4x_2$
2. $\hat{y} = 95 + 1.5x_1 + 3x_2 + 2x_1x_2$

Both models have been built over the range $0.5 \le x_2 \le 10$.

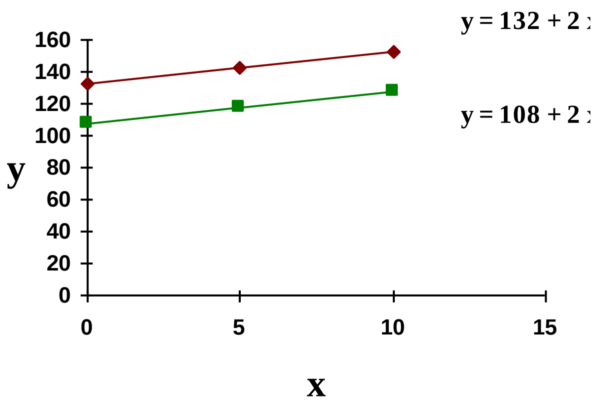
(a) What is the predicted value of conversion when x_2 = 2?
Repeat this calculation for x_2 = 8. Draw a graph of the predicted values for both conversion models. Comment on the effect of the interaction term in model 2.
(b) Find the expected change in the mean conversion for a unit change in temperature x_1 for model 1 when x_2 = 5.
Does this quantity depend on the specific value of reaction time selected? Why?

(c) Find the expected change in the mean conversion for a unit change in temperature x_1 for model 2 when $x_2 = 5$. Repeat this calculation for $x_2 = 2$ and $x_2 = 8$. Does the result depend on the value selected for x_2? Why?

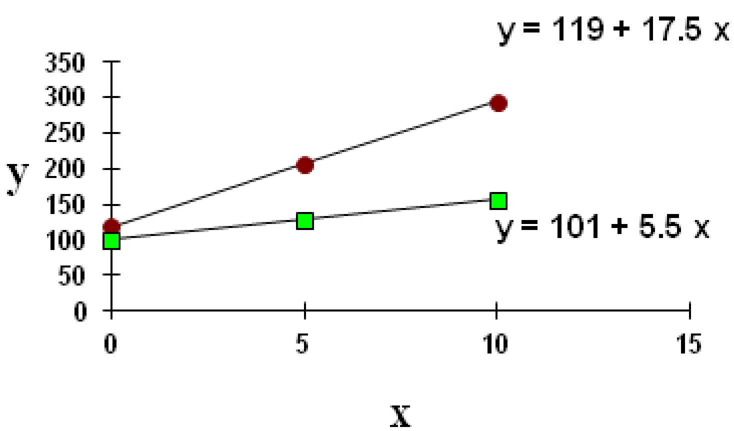
(a)

	Model 1	Model 2
$x_2 = 2$	$\hat{y} = 100 + 2x_1 + 8$	$\hat{y} = 95 + 1.5x_1 + 3(2) + 4x_1$
	$\hat{y} = 108 + 2x_1$	$\hat{y} = 101 + 5.5x_1$
$x_2 = 8$	$\hat{y} = 100 + 2x_1 + 4(8)$	$\hat{y} = 95 + 1.5x_1 + 3(8) + 16x_1$
	$\hat{y} = 132 + 2x_1$	$\hat{y} = 119 + 17.5x_1$

MODEL 1



MODEL 2



The interaction term in model 2 affects the slope of the regression equation. That is, it modifies the amount of change per unit of x_1 on $\hat{y}$.

(b) $x_2 = 5$ $\quad \hat{y} = 100 + 2x_1 + 4(5)$

$\hat{y} = 120 + 2x_1$

Then, 2 is the expected change on $\hat{y}$ per unit of x_1.

No, it does not depend on the value of x_2, because there is no relationship or interaction between these two variables in model 1.

(c)

	$x_2 = 5$	$x_2 = 2$	$x_2 = 8$
	$\hat{y} = 95 + 1.5x_1 + 3(5) + 2x_1(5)$ $\hat{y} = 110 + 11.5x_1$	$\hat{y} = 101 + 5.5x_1$	$\hat{y} = 119 + 17.5x_1$
Change per unit of x_1	11.5	5.5	17.5

Yes, the result does depend on the value of x_2, because x_2 interacts with x_1.

12-8 You have fit a multiple linear regression model and the $(X'X)^{-1}$ matrix is:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.893758 & -0.0282448 & -0.0175641 \\ -0.028245 & 0.013329 & 0.0001547 \\ -0.017564 & 0.0001547 & 0.0009108 \end{bmatrix}$$

(a) How many regressor variables are in this model?
(b) If the error sum of squares is 307 and there are 15 observations, what is the estimate of σ^2?
(c) What is the standard error of the regression coefficient $\hat{\beta}_1$?

(a) There are two regressor variables in this model based on the size of the $(X'X)^{-1}$ matrix.

(b) The estimate of σ^2 is the $MS_{Residual}$. The $MS_{Residual} = \dfrac{SS_{Residual}}{DF} = \dfrac{307}{14-2} = 25.583$

(c) Standard error of $\hat{\beta}_1 = \sqrt{\hat{\sigma}^2 C_{jj}} = \sqrt{(25.583)(0.0013329)} = 0.1847$

12-9 The data from a patient satisfaction survey in a hospital are in Table E12-1.

TABLE • E12-1 Patient Satisfaction Data

Observation	Age	Severity	Surg-Med	Anxiety	Satisfaction
1	55	50	0	2.1	68
2	46	24	1	2.8	77
3	30	46	1	3.3	96
4	35	48	1	4.5	80
5	59	58	0	2.0	43
6	61	60	0	5.1	44
7	74	65	1	5.5	26
8	38	42	1	3.2	88
9	27	42	0	3.1	75
10	51	50	1	2.4	57
11	53	38	1	2.2	56
12	41	30	0	2.1	88
13	37	31	0	1.9	88
14	24	34	0	3.1	102
15	42	30	0	3.0	88
16	50	48	1	4.2	70
17	58	61	1	4.6	52
18	60	71	1	5.3	43
19	62	62	0	7.2	46
20	68	38	0	7.8	56
21	70	41	1	7.0	59
22	79	66	1	6.2	26
23	63	31	1	4.1	52
24	39	42	0	3.5	83
25	49	40	1	2.1	75

The regressor variables are the patient's age, an illness severity index (higher values indicate greater severity), an indicator variable denoting whether the patient is a medical patient (0) or a surgical patient (1), and an anxiety index (higher values indicate greater anxiety).

(a) Fit a multiple linear regression model to the satisfaction response using age, illness severity, and the anxiety index as the regressors.
(b) Estimate σ^2.
(c) Find the standard errors of the regression coefficients.
(d) Are all of the model parameters estimated with nearly the same precision? Why or why not?

(a) The results from computer software follow. The model can be expressed as

```
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety
```

(b) $\hat{\sigma}^2 = \dfrac{\sum_{t=1}^{n} e_t^2}{n-p} = \dfrac{SS_E}{n-p} = \dfrac{1039.9}{21} = 49.5$

(c) $\text{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 C$, $se(\hat{\beta}) = \sqrt{\hat{\sigma}^2 C_{jj}} = \begin{bmatrix} 5.9 \\ 0.13 \\ 0.13 \\ 1.06 \end{bmatrix}$ from the Minitab output.

(d) Because the regression coefficients have different standard errors the parameters estimators do not have similar precision of estimation.

Regression Analysis: Satisfaction versus Age, Severity, Anxiety

```
The regression equation is
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety

Predictor     Coef  SE Coef      T      P
Constant   143.895    5.898  24.40  0.000
Age        -1.1135   0.1326  -8.40  0.000
Severity   -0.5849   0.1320  -4.43  0.000
Anxiety      1.296    1.056   1.23  0.233

S = 7.03710   R-Sq = 90.4%   R-Sq(adj) = 89.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3   9738.3  3246.1  65.55  0.000
Residual Error  21   1039.9    49.5
Total           24  10778.2

Source    DF  Seq SS
Age        1  8756.7
Severity   1   907.0
Anxiety    1    74.6

Unusual Observations

Obs   Age  Satisfaction    Fit  SE Fit  Residual  St Resid
  9  27.0         75.00  93.28    2.98    -18.28     -2.87R
```

12-10 The electric power consumed each month by a chemical plant is thought to be related to the average ambient temperature (x_1), the number of days in the month (x_2), the average product purity (x_3), and the tons of product produced (x_4). The past year's historical data are available and are presented in Table E12-2.

(a) Fit a multiple linear regression model to these data.
(b) Estimate σ^2 .
(c) Compute the standard errors of the regression coefficients.
Are all of the model parameters estimated with the same precision? Why or why not?
(d) Predict power consumption for a month in which x_1 = 75°F, x_2 = 24 days, x_3 = 90%, and x_4 = 98 tons.

TABLE • E12-2 Power Consumption Data

y	x_1	x_2	x_3	x_4
240	25	24	91	100
236	31	21	90	95
270	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

```
Predictor          Coef     SE Coef         T        P
Constant         -123.1       157.3     -0.78    0.459
X1               0.7573      0.2791      2.71    0.030
X2                7.519       4.010      1.87    0.103
X3                2.483       1.809      1.37    0.212
X4              -0.4811      0.5552     -0.87    0.415

S = 11.79        R-Sq = 85.2%      R-Sq(adj) = 76.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4      5600.5      1400.1     10.08    0.005
Residual Error     7       972.5       138.9
Total             11      6572.9
```

(a)

$$\hat{y} = -123.1 + 0.7573x_1 + 7.519x_2 + 2.483x_3 - 0.4811x_4$$

(b)

$$\hat{\sigma}^2 = 139.00$$

(c)

$se(\hat{\beta}_0) = 157.3$, $se(\hat{\beta}_1) = 0.2791$, $se(\hat{\beta}_2) = 4.010$, $se(\hat{\beta}_3) = 1.809$, and $se(\hat{\beta}_4) = 0.5552$

Because the regression coefficients have different standard errors the parameters estimators do not have similar precision of estimation.

(a) $\hat{y} = -123.1 + 0.7573(75) + 7.519(24) + 2.483(90) - 0.4811(98) = 290.476$

12-11 Table E12-3 provides the highway gasoline mileage test results for 2005 model year vehicles from DaimlerChrysler. The full table of data (available on the book's Web site) contains

TABLE • E12-3 DaimlerChrysler Fuel Economy and Emissions

mfr	carline	car/truck	cid	rhp	trns	drv	od	etw	cmp	axle	n/v	a/c	hc	co	co2	mpg
20	300C/SRT-8	C	215	253	L5	4	2	4500	9.9	3.07	30.9	Y	0.011	0.09	288	30.8
20	CARAVAN 2WD	T	201	180	L4	F	2	4500	9.3	2.49	32.3	Y	0.014	0.11	274	32.5
20	CROSSFIRE ROADSTER	C	196	168	L5	R	2	3375	10	3.27	37.1	Y	0.001	0.02	250	35.4
20	DAKOTA PICKUP 2WD	T	226	210	L4	R	2	4500	9.2	3.55	29.6	Y	0.012	0.04	316	28.1
20	DAKOTA PICKUP 4WD	T	226	210	L4	4	2	5000	9.2	3.55	29.6	Y	0.011	0.05	365	24.4
20	DURANGO 2WD	T	348	345	L5	R	2	5250	8.6	3.55	27.2	Y	0.023	0.15	367	24.1
20	GRAND CHEROKEE 2WD	T	226	210	L4	R	2	4500	9.2	3.07	30.4	Y	0.006	0.09	312	28.5
20	GRAND CHEROKEE 4WD	T	348	230	L5	4	2	5000	9	3.07	24.7	Y	0.008	0.11	369	24.2
20	LIBERTY/CHEROKEE 2WD	T	148	150	M6	R	2	4000	9.5	4.1	41	Y	0.004	0.41	270	32.8
20	LIBERTY/CHEROKEE 4WD	T	226	210	L4	4	2	4250	9.2	3.73	31.2	Y	0.003	0.04	317	28
20	NEON/SRT-4/SX 2.0	C	122	132	L4	F	2	3000	9.8	2.69	39.2	Y	0.003	0.16	214	41.3
20	PACIFICA 2WD	T	215	249	L4	F	2	4750	9.9	2.95	35.3	Y	0.022	0.01	295	30
20	PACIFICA AWD	T	215	249	L4	4	2	5000	9.9	2.95	35.3	Y	0.024	0.05	314	28.2
20	PT CRUISER	T	148	220	L4	F	2	3625	9.5	2.69	37.3	Y	0.002	0.03	260	34.1
20	RAM 1500 PICKUP 2WD	T	500	500	M6	R	2	5250	9.6	4.1	22.3	Y	0.01	0.1	474	18.7
20	RAM 1500 PICKUP 4WD	T	348	345	L5	4	2	6000	8.6	3.92	29	Y	0	0	0	20.3
20	SEBRING 4-DR	C	165	200	L4	F	2	3625	9.7	2.69	36.8	Y	0.011	0.12	252	35.1
20	STRATUS 4-DR	C	148	167	L4	F	2	3500	9.5	2.69	36.8	Y	0.002	0.06	233	37.9
20	TOWN & COUNTRY 2WD	T	148	150	L4	F	2	4250	9.4	2.69	34.9	Y	0	0.09	262	33.8
20	VIPER CONVERTIBLE	C	500	501	M6	R	2	3750	9.6	3.07	19.4	Y	0.007	0.05	342	25.9
20	WRANGLER/TJ 4WD	T	148	150	M6	4	2	3625	9.5	3.73	40.1	Y	0.004	0.43	337	26.4

mfr-mfr code
carline-car line name (test vehicle model name)
car/truck-'C' for passenger vehicle and 'T' for truck
cid-cubic inch displacement of test vehicle
rhp-rated horsepower
trns-transmission code
drv-drive system code
od-overdrive code
etw-equivalent test weight
cmp-compression ratio
axle-axle ratio
n/v-n/v ratio (engine speed versus vehicle speed at 50 mph)
a/c-indicates air conditioning simulation
hc-HC(hydrocarbon emissions) Test level composite results
co-CO(carbon monoxide emissions) Test level composite results
co2-CO2(carbon dioxide emissions) Test level composite results
mpg-mpg(fuel economy, miles per gallon)

the same data for 2005 models from over 250 vehicles from many manufacturers (Environmental Protection Agency Web site www.epa.gov/ otaq/cert/mpg/testcars/database).

(a) Fit a multiple linear regression model to these data to estimate gasoline mileage that uses the following regressors: *cid, rhp, etw, cmp, axle, n/v*
(b) Estimate σ^2 and the standard errors of the regression coefficients.
(c) Predict the gasoline mileage for the first vehicle in the table.

```
The regression equation is
mpg = 49.9 - 0.0104 cid - 0.0012 rhp - 0.00324 etw + 0.29 cmp - 3.86 axle
      + 0.190 n/v

Predictor           Coef     SE Coef       T       P
Constant           49.90       19.67    2.54   0.024
cid             -0.01045     0.02338   -0.45   0.662
rhp             -0.00120     0.01631   -0.07   0.942
etw           -0.0032364   0.0009459   -3.42   0.004
cmp                0.292       1.765    0.17   0.871
axle              -3.855       1.329   -2.90   0.012
n/v               0.1897      0.2730    0.69   0.498

S = 2.22830   R-Sq = 89.3%   R-Sq(adj) = 84.8%

Analysis of Variance

Source            DF        SS       MS       F      P
```

```
Regression        6   581.898   96.983   19.53   0.000
Residual Error   14    69.514    4.965
Total            20   651.412
```

(a) $\hat{y} = 49.90 - 0.01045x_1 - 0.0012x_2 - 0.00324x_3 + 0.292x_4 - 3.855x_5 + 0.1897x_6$

where $x_1 = cid \quad x_2 = rhp \quad x_3 = etw \quad x_4 = cmp \quad x_5 = axle \quad x_6 = n/v$

(b) $\hat{\sigma}^2 = 4.965$

$se(\hat{\beta}_0) = 19.67$, $se(\hat{\beta}_1) = 0.02338$, $se(\hat{\beta}_2) = 0.01631$, $se(\hat{\beta}_3) = 0.0009459$,
$se(\hat{\beta}_4) = 1.765$, $se(\hat{\beta}_5) = 1.329$ and $se(\hat{\beta}_6) = 0.273$

(c)
$\hat{y} = 49.90 - 0.01045(215) - 0.0012(253) - 0.0032(4500) + 0.292(9.9) - 3.855(3.07) + 0.1897(30.9)$
= 29.867

12-12 The pull strength of a wire bond is an important characteristic. Table E12-4 gives information on pull strength (y), die height (x_1), post height (x_2), loop height (x_3), wire length (x_4), bond width on the die (x_5), and bond width on the post (x_6).

(a) Fit a multiple linear regression model using x_2, x_3, x_4, and x_5 as the regressors.
(b) Estimate σ^2 .
(c) Find the $se(\hat{\beta}_j)$. How precisely are the regression coefficients estimated in your opinion?
(d) Use the model from part (a) to predict pull strength when x_2 = 20, x_3 = 30, x_4 = 90, and x_5 = 2.0.

TABLE • E12-4 Wire Bond Data

y	x_1	x_2	x_3	x_4	x_5	x_6
8.0	5.2	19.6	29.6	94.9	2.1	2.3
8.3	5.2	19.8	32.4	89.7	2.1	1.8
8.5	5.8	19.6	31.0	96.2	2.0	2.0
8.8	6.4	19.4	32.4	95.6	2.2	2.1
9.0	5.8	18.6	28.6	86.5	2.0	1.8
9.3	5.2	18.8	30.6	84.5	2.1	2.1
9.3	5.6	20.4	32.4	88.8	2.2	1.9
9.5	6.0	19.0	32.6	85.7	2.1	1.9
9.8	5.2	20.8	32.2	93.6	2.3	2.1
10.0	5.8	19.9	31.8	86.0	2.1	1.8
10.3	6.4	18.0	32.6	87.1	2.0	1.6
10.5	6.0	20.6	33.4	93.1	2.1	2.1
10.8	6.2	20.2	31.8	83.4	2.2	2.1
11.0	6.2	20.2	32.4	94.5	2.1	1.9
11.3	6.2	19.2	31.4	83.4	1.9	1.8
11.5	5.6	17.0	33.2	85.2	2.1	2.1
11.8	6.0	19.8	35.4	84.1	2.0	1.8
12.3	5.8	18.8	34.0	86.9	2.1	1.8
12.5	5.6	18.6	34.2	83.0	1.9	2.0

```
The regression equation is
y = 7.46 - 0.030 x2 + 0.521 x3 - 0.102 x4 - 2.16 x5
Predictor        Coef        StDev          T        P
Constant        7.458        7.226       1.03    0.320
x2            -0.0297       0.2633      -0.11    0.912
```

```
x3             0.5205       0.1359       3.83     0.002
x4           -0.10180      0.05339      -1.91     0.077
x5             -2.161        2.395      -0.90     0.382
S = 0.8827      R-Sq = 67.2%     R-Sq(adj) = 57.8%
Analysis of Variance
Source         DF          SS          MS         F        P
Regression      4     22.3119      5.5780      7.16    0.002
Error          14     10.9091      0.7792
Total          18     33.2211
```

(a) $\hat{y} = 7.4578 - 0.0297x_2 + 0.5205x_3 - 0.1018x_4 - 2.1606x_5$

(b) $\hat{\sigma}^2 = .7792$

(c) $se(\hat{\beta}_0) = 7.226$, $se(\hat{\beta}_2) = .2633$, $se(\hat{\beta}_3) = .1359$, $se(\hat{\beta}_4) = .05339$ and $se(\hat{\beta}_5) = 2.395$

(d) $\hat{y} = 7.4578 - 0.0297(20) + 0.5205(30) - 0.1018(90) - 2.1606(2.0)$ $\quad \hat{y} = 8.996$

12-13 An engineer at a semiconductor company wants to model the relationship between the device HFE (y) and three parameters: Emitter-RS (x_1), Base-RS (x_2), and Emitter-to- Base RS (x_3). The data are shown in the Table E12-5.

(a) Fit a multiple linear regression model to the data.
(b) Estimate σ^2.
(c) Find the standard errors $se(\hat{\beta}_j)$. Are all of the model parameters estimated with the same precision? Justify your answer.
(d) Predict HFE when x_1 = 14.5, x_2 = 220, and x_3 =5.0.

TABLE • E12-5 Semiconductor Data

x_1 Emitter-RS	x_2 Base-RS	x_3 E-B-RS	y HFE-1M-5V
14.620	226.00	7.000	128.40
15.630	220.00	3.375	52.62
14.620	217.40	6.375	113.90
15.000	220.00	6.000	98.01
14.500	226.50	7.625	139.90
15.250	224.10	6.000	102.60
16.120	220.50	3.375	48.14
15.130	223.50	6.125	109.60
15.500	217.60	5.000	82.68b
15.130	228.50	6.625	112.60
15.500	230.20	5.750	97.52
16.120	226.50	3.750	59.06
15.130	226.60	6.125	111.80
15.630	225.60	5.375	89.09
15.380	229.70	5.875	101.00
14.380	234.00	8.875	171.90
15.500	230.00	4.000	66.80
14.250	224.30	8.000	157.10
14.500	240.50	10.870	208.40
14.620	223.70	7.375	133.40

Regression Analysis: Ex12-9y versus Ex12-9x1, Ex12-9x2, Ex12-9x3

```
The regression equation is
Ex12-9y = 47.8 - 9.60 Ex12-9x1 + 0.415 Ex12-9x2 + 18.3 Ex12-9x3
```

```
Predictor     Coef   SE Coef       T       P
Constant     47.82     49.94    0.96   0.353
Ex12-9x1    -9.604     3.723   -2.58   0.020
Ex12-9x2    0.4152    0.2261    1.84   0.085
Ex12-9x3    18.294     1.323   13.82   0.000


S = 3.50508   R-Sq = 99.4%   R-Sq(adj) = 99.2%


Analysis of Variance

Source           DF     SS     MS       F      P
Regression        3  30529  10176  828.31  0.000
Residual Error   16    197     12
Total            19  30725
```

(a) y hat = 47.8 - 9.60x1 + 0.415x2 + 18.3x3

(b) $\hat{\sigma}^2 = 12$

(c) The estimated standard errors of the coefficient estimators are provided in the above table (SE Coef). Because the regression coefficients have different standard errors the parameters estimators do not have similar precision of estimation.

(d) y hat = 47.8 - 9.60(14.5) + 0.415(220) + 18.3(5) = 91.372

12-14 Heat treating is often used to carburize metal parts such as gears. The thickness of the carburized layer is considered a crucial feature of the gear and contributes to the overall reliability of the part. Because of the critical nature of this feature, two different lab tests are performed on each furnace load. One test is run on a sample pin that accompanies each load. The other test is a destructive test that cross-sections an actual part. This test involves running a carbon analysis on the surface of both the gear pitch (top of the gear tooth) and the gear root (between the gear teeth). Table E12-6 shows the results of the pitch carbon analysis test for 32 parts.

TABLE • E12-6 Heat Treating Test

TEMP	SOAKTIME	SOAKPCT	DIFFTIME	DIFFPCT	PITCH
1650	0.58	1.10	0.25	0.90	0.013
1650	0.66	1.10	0.33	0.90	0.016
1650	0.66	1.10	0.33	0.90	0.015
1650	0.66	1.10	0.33	0.95	0.016
1600	0.66	1.15	0.33	1.00	0.015
1600	0.66	1.15	0.33	1.00	0.016
1650	1.00	1.10	0.50	0.80	0.014
1650	1.17	1.10	0.58	0.80	0.021
1650	1.17	1.10	0.58	0.80	0.018
1650	1.17	1.10	0.58	0.80	0.019
1650	1.17	1.10	0.58	0.90	0.021
1650	1.17	1.10	0.58	0.90	0.019

TEMP	SOAKTIME	SOAKPCT	DIFFTIME	DIFFPCT	PITCH
1650	1.17	1.15	0.58	0.90	0.021
1650	1.20	1.15	1.10	0.80	0.025
1650	2.00	1.15	1.00	0.80	0.025
1650	2.00	1.10	1.10	0.80	0.026
1650	2.20	1.10	1.10	0.80	0.024
1650	2.20	1.10	1.10	0.80	0.025
1650	2.20	1.15	1.10	0.80	0.024
1650	2.20	1.10	1.10	0.90	0.025
1650	2.20	1.10	1.10	0.90	0.027
1650	2.20	1.10	1.50	0.90	0.026
1650	3.00	1.15	1.50	0.80	0.029
1650	3.00	1.10	1.50	0.70	0.030
1650	3.00	1.10	1.50	0.75	0.028
1650	3.00	1.15	1.66	0.85	0.032
1650	3.33	1.10	1.50	0.80	0.033
1700	4.00	1.10	1.50	0.70	0.039
1650	4.00	1.10	1.50	0.70	0.040
1650	4.00	1.15	1.50	0.85	0.035
1700	12.50	1.00	1.50	0.70	0.056
1700	18.50	1.00	1.50	0.70	0.068

The regressors are furnace temperature (TEMP), carbon concentration and duration of the carburizing cycle (SOAKPCT,SOAKTIME), and carbon concentration and duration of the diffuse cycle (DIFFPCT, DIFFTIME).

(a) Fit a linear regression model relating the results of the pitch carbon analysis test (PITCH) to the five regressor variables.
(b) Estimate σ^2.
(c) Find the standard errors $se(\hat{\beta}_j)$
(d) Use the model in part (a) to predict PITCH when TEMP = 1650, SOAKTIME = 1.00, SOAKPCT = 1.10, DIFFTIME = 1.00, and DIFFPCT = 0.80.

```
Predictor         Coef      SE Coef          T        P
Constant      -0.03023    0.06178        -0.49    0.629
temp        0.00002856    0.00003437      0.83    0.414
soaktime     0.0023182    0.0001737      13.35    0.000
soakpct      -0.003029    0.005844       -0.52    0.609
difftime      0.008476    0.001218        6.96    0.000
diffpct      -0.002363    0.008078       -0.29    0.772

S = 0.002296    R-Sq = 96.8%     R-Sq(adj) = 96.2%

Analysis of Variance

Source              DF           SS          MS         F        P
Regression           5   0.00418939  0.00083788    158.92    0.000
Residual Error      26   0.00013708  0.00000527
Total               31   0.00432647
```

(a) $\hat{y} = -0.03023 + 0.000029x_1 + 0.002318x_2 - 0.003029x_3 + 0.008476x_4 - 0.002363x_5$

where $x_1 = TEMP$ $x_2 = SOAKTIME$ $x_3 = SOAKPCT$ $x_4 = DFTIME$ $x_5 = DIFFPCT$

(b) $\hat{\sigma}^2 = 5.27x10^{-6}$

(c) The standard errors are listed under the StDev column above.

(d)

$$\hat{y} = -0.03023 + 0.000029(1650) + 0.002318(1) - 0.003029(1.1)$$
$$+ 0.008476(1) - 0.002363(0.80)$$

$$\hat{y} = 0.02247$$

12-15 An article in *Electronic Packaging and Production* (2002, Vol. 42) considered the effect of X-ray inspection of integrated circuits. The rads (radiation dose) were studied as a function of current (in milliamps) and exposure time (in minutes). The data are in Table E12-7.

TABLE • E12-7 X-ray Inspection Data

Rads	mAmps	Exposure Time
7.4	10	0.25
14.8	10	0.5
29.6	10	1
59.2	10	2
88.8	10	3
296	10	10
444	10	15
592	10	20
11.1	15	0.25
22.2	15	0.5
44.4	15	1
88.8	15	2
133.2	15	3
444	15	10
666	15	15
888	15	20
14.8	20	0.25
29.6	20	0.5
59.2	20	1
118.4	20	2
177.6	20	3
592	20	10
888	20	15
1184	20	20
22.2	30	0.25
44.4	30	0.5
88.8	30	1
177.6	30	2
266.4	30	3
888	30	10
1332	30	15
1776	30	20
29.6	40	0.25
59.2	40	0.5
118.4	40	1
236.8	40	2
355.2	40	3
1184	40	10
1776	40	15
2368	40	20

(a) Fit a multiple linear regression model to these data with rads as the response.
(b) Estimate σ^2 and the standard errors of the regression coefficients.
(c) Use the model to predict rads when the current is 15 milliamps and the exposure time is 5 seconds.

```
The regression equation is
rads = - 440 + 19.1 mAmps + 68.1 exposure time
```

```
Predictor         Coef   SE Coef      T      P
Constant        -440.39    94.20  -4.68  0.000
mAmps            19.147    3.460   5.53  0.000
exposure time    68.080    5.241  12.99  0.000

S = 235.718   R-Sq = 84.3%   R-Sq(adj) = 83.5%

Analysis of Variance

Source          DF        SS       MS      F      P
Regression       2  11076473  5538237  99.67  0.000
Residual Error  37   2055837    55563
Total           39  13132310
```

(a) $\hat{y} = -440.39 + 19.147x_1 + 68.080x_2$

where $x_1 = mAmps \quad x_2 = ExposureTime$

(b) $\hat{\sigma}^2 = 55563$

$se(\hat{\beta}_0) = 94.20$, $se(\hat{\beta}_1) = 3.460$, and $se(\hat{\beta}_2) = 5.241$

(c) $\hat{y} = -440.93 + 19.147(15) + 68.080(5) = 186.675$

12-16 An article in *Cancer Epidemiology, Biomarkers and Prevention* (1996, Vol. 5, pp. 849–852) reported on a pilot study to assess the use of toenail arsenic concentrations as an indicator of ingestion of arsenic-containing water. Twenty-one participants were interviewed regarding use of their private (unregulated) wells for drinking and cooking, and each provided a sample of water and toenail clippings. Table E12-8 showed the data of age (years), sex of person (1 = male, 2 = female), proportion of times household well used for drinking ($1 \le 1/4$, $2 = 1/4$, $3 = 1/2$, $4 = 3/4$, $5 \ge 3/4$), proportion of times household well used for cooking ($1 \le 1/4$, $2 = 1/4$, $3 = 1/2$, $4 = 3/4$, $5 \ge 3/4$), arsenic in water (ppm), and arsenic in toenails (ppm) respectively.

(a) Fit a multiple linear regression model using arsenic concentration in nails as the response and age, drink use, cook use, and arsenic in the water as the regressors.
(b) Estimate σ^2 and the standard errors of the regression coefficients.
(c) Use the model to predict the arsenic in nails when the age is 30, the drink use is category 5, the cook use is category 5, and arsenic in the water is 0.135 ppm.

TABLE • E12-8 Arsenic Data

Age	Sex	Drink Use	Cook Use	Arsenic Water	Arsenic Nails
44	2	5	5	0.00087	0.119
45	2	4	5	0.00021	0.118
44	1	5	5	0	0.099
66	2	3	5	0.00115	0.118
37	1	2	5	0	0.277
45	2	5	5	0	0.358
47	1	5	5	0.00013	0.08
38	2	4	5	0.00069	0.158
41	2	3	2	0.00039	0.31
49	2	4	5	0	0.105
72	2	5	5	0	0.073
45	2	1	5	0.046	0.832
53	1	5	5	0.0194	0.517
86	2	5	5	0.137	2.252
8	2	5	5	0.0214	0.851
32	2	5	5	0.0175	0.269
44	1	5	5	0.0764	0.433
63	2	5	5	0	0.141
42	1	5	5	0.0165	0.275
62	1	5	5	0.00012	0.135
36	1	5	5	0.0041	0.175

```
The regression equation is
```

```
ARSNAILS = 0.488 - 0.00077 AGE - 0.0227 DRINKUSE - 0.0415 COOKUSE
           + 13.2 ARSWATER

Predictor        Coef   SE Coef      T      P
Constant       0.4875    0.4272   1.14  0.271
AGE         -0.000767  0.003508  -0.22  0.830
DRINKUSE     -0.02274   0.04747  -0.48  0.638
COOKUSE      -0.04150   0.08408  -0.49  0.628
ARSWATER       13.240     1.679   7.89  0.000

S = 0.236010   R-Sq = 81.2%   R-Sq(adj) = 76.5%

Analysis of Variance

Source          DF        SS       MS      F      P
Regression       4   3.84906  0.96227  17.28  0.000
Residual Error  16   0.89121  0.05570
Total           20   4.74028
```

(a) $\hat{y} = 0.4875 - 0.000767x_1 - 0.02274x_2 - 0.04150x_3 + 13.240x_4$

where $x_1 = AGE \quad x_2 = DrinkUse \quad x_3 = CookUse \quad x_4 = ARSWater$

(b) $\hat{\sigma}^2 = 0.05570$

$se(\hat{\beta}_0) = 0.4272$, $se(\hat{\beta}_1) = 0.003508$, $se(\hat{\beta}_2) = 0.04747$, $se(\hat{\beta}_3) = 0.08408$, and $se(\hat{\beta}_4) = 1.679$

(c) $\hat{y} = 0.4875 - 0.000767(30) - 0.02274(5) - 0.04150(5) + 13.240(0.135) = 1.9307$

12-17 An article in *IEEE Transactions on Instrumentation and Measurement* (2001, Vol. 50, pp. 2033–2040) reported on a study that had analyzed powdered mixtures of coal and limestone for permittivity. The errors in the density measurement was the response. The data are reported in Table E12-9.

TABLE • E12-9 Density Data

Density	Dielectric Constant	Loss Factor
0.749	2.05	0.016
0.798	2.15	0.02
0.849	2.25	0.022
0.877	2.3	0.023
0.929	2.4	0.026
0.963	2.47	0.028
0.997	2.54	0.031
1.046	2.64	0.034
1.133	2.85	0.039
1.17	2.94	0.042
1.215	3.05	0.045

(a) Fit a multiple linear regression model to these data with the density as the response.
(b) Estimate σ^2 and the standard errors of the regression coefficients.
(c) Use the model to predict the density when the dielectric constant is 2.5 and the loss factor is 0.03.

```
The regression equation is
density = - 0.110 + 0.407 dielectric constant + 2.11 loss factor

Predictor                    Coef   SE Coef       T       P
```

```
Constant              -0.1105   0.2501  -0.44  0.670
dielectric constant    0.4072   0.1682   2.42  0.042
loss factor             2.108    5.834   0.36  0.727

S = 0.00883422   R-Sq = 99.7%   R-Sq(adj) = 99.7%

Analysis of Variance

Source          DF       SS       MS        F      P
Regression       2  0.23563  0.11782  1509.64  0.000
Residual Error   8  0.00062  0.00008
Total           10  0.23626
```

(a) $\hat{y} = -0.1105 + 0.4072x_1 + 2.108x_2$

where $x_1 = Dielectric\,Const \quad x_2 = LossFactor$

(b) $\hat{\sigma}^2 = 0.00008$

$se(\hat{\beta}_0) = 0.2501$, $se(\hat{\beta}_1) = 0.1682$, and $se(\hat{\beta}_2) = 5.834$

(c) $\hat{y} = -0.1105 + 0.4072(2.5) + 2.108(0.03) = 0.97074$

12-18 An article in *Biotechnology Progress* (2001, Vol. 17, pp. 366–368) reported on an experiment to investigate and optimize nisin extraction in aqueous two-phase systems (ATPS). The nisin recovery was the dependent variable (y). The two regressor variables were concentration (%) of PEG 4000 (denoted as x_1 and concentration (%) of Na_2SO_4 (denoted as x_2). The data are in Table E12-10.

TABLE • E12-10 Nisin Extraction Data

x_1	x_2	y
13	11	62.8739
15	11	76.1328
13	13	87.4667
15	13	102.3236
14	12	76.1872
14	12	77.5287
14	12	76.7824
14	12	77.4381
14	12	78.7417

(a) Fit a multiple linear regression model to these data.
(b) Estimate σ^2 and the standard errors of the regression coefficients.
(c) Use the model to predict the nisin recovery when $x_1 = 14.5$ and $x_2 = 12.5$.

```
The regression equation is
y = - 171 + 7.03 x1 + 12.7 x2

Predictor      Coef  SE Coef      T      P
Constant    -171.26    28.40  -6.03  0.001
x1            7.029    1.539   4.57  0.004
x2           12.696    1.539   8.25  0.000

S = 3.07827   R-Sq = 93.7%   R-Sq(adj) = 91.6%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  842.37  421.18  44.45  0.000
```

```
Residual Error   6   56.85    9.48
Total            8  899.22
```

(a) $\hat{y} = -171 + 7.03x_1 + 12.7x_2$

(b) $\hat{\sigma}^2 = 9.48$

$se(\hat{\beta}_0) = 28.40$, $se(\hat{\beta}_1) = 1.539$, and $se(\hat{\beta}_2) = 1.539$

(c) $\hat{y} = -171 + 7.03(14.5) + 12.7(12.5)$

= 89.685

12-19 An article in *Optical Engineering* ["Operating Curve Extraction of a Correlator's Filter" (2004, Vol. 43, pp. 2775–2779)] reported on the use of an optical correlator to perform an experiment by varying brightness and contrast. The resulting modulation is characterized by the useful range of gray levels. The data follow:

Brightness (%):	54	61	65	100	100	100	50	57	54
Contrast (%):	56	80	70	50	65	80	25	35	26
Useful range (ng):	96	50	50	112	96	80	155	144	255

(a) Fit a multiple linear regression model to these data.
(b) Estimate σ^2.
(c) Compute the standard errors of the regression coefficients.
(d) Predict the useful range when brightness = 80 and contrast= 75.

```
The regression equation is
Useful range (ng) = 239 + 0.334 Brightness (%) - 2.72 Contrast (%)


Predictor          Coef  SE Coef      T      P
Constant         238.56    45.23   5.27  0.002
Brightness (%)   0.3339   0.6763   0.49  0.639
Contrast (%)    -2.7167   0.6887  -3.94  0.008


S = 36.3493   R-Sq = 75.6%   R-Sq(adj) = 67.4%

Analysis of Variance

Source          DF     SS     MS     F      P
Regression       2  24518  12259  9.28  0.015
Residual Error   6   7928   1321
Total            8  32446
```

(a) $\hat{y} = 238.56 + 0.3339x_1 - 2.7167x_2$

where $x_1 = \%Brightness \quad x_2 = \%Contrast$

(b) $\hat{\sigma}^2 = 1321$

(c) $se(\hat{\beta}_0) = 45.23$, $se(\hat{\beta}_1) = 0.6763$, and $se(\hat{\beta}_2) = 0.6887$

(d) $\hat{y} = 238.56 + 0.3339(80) - 2.7167(75) = 61.5195$

12-20 An article in *Technometrics* (1974, Vol. 16, pp. 523–531) considered the following stack-loss data from a plant oxidizing ammonia to nitric acid. Twenty-one daily responses of stack loss (the amount of ammonia escaping) were measured with air flow x_1, temperature x_2, and acid concentration x_3.

y = 42, 37, 37, 28, 18, 18, 19, 20, 15, 14, 14, 13, 11, 12, 8, 7, 8, 8, 9, 15, 15
x_1 = 80, 80, 75, 62, 62, 62, 62, 62, 58, 58, 58, 58, 58, 58, 50, 50, 50, 50, 50, 56, 70
x_2 = 27, 27, 25, 24, 22, 23, 24, 24, 23, 18, 18, 17, 18, 19, 18, 18, 19, 19, 20, 20, 20
x_3 = 89, 88, 90, 87, 87, 87, 93, 93, 87, 80, 89, 88, 82, 93, 89, 86, 72, 79, 80, 82, 91

(a) Fit a linear regression model relating the results of the stack loss to the three regressor variables.
(b) Estimate σ^2.
(c) Find the standard error $se(\hat{\beta}_j)$.
(d) Use the model in part (a) to predict stack loss when $x_1 = 60$, $x_2 = 26$, and $x_3 = 85$.

```
The regression equation is
Stack Loss(y) = - 39.9 + 0.716 X1 + 1.30 X2 - 0.152 X3

Predictor     Coef  SE Coef      T      P
Constant    -39.92    11.90  -3.36  0.004
X1          0.7156   0.1349   5.31  0.000
X2          1.2953   0.3680   3.52  0.003
X3         -0.1521   0.1563  -0.97  0.344

S = 3.24336   R-Sq = 91.4%   R-Sq(adj) = 89.8%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3  1890.41  630.14  59.90  0.000
Residual Error  17   178.83   10.52
Total           20  2069.24
```

(a) $\hat{y} = -39.92 + 0.7156x_1 + 1.2953x_2 - 0.1521x_3$

(b) $\hat{\sigma}^2 = 10.52$

(c) $se(\hat{\beta}_0) = 11.90$, $se(\hat{\beta}_1) = 0.1349$, $se(\hat{\beta}_2) = 0.3680$, and $se(\hat{\beta}_3) = 0.1563$

(d) $\hat{y} = -39.92 + 0.7156(60) + 1.2953(26) - 0.1521(85) = 23.7653$

12-21 Table E12-11 presents quarterback ratings for the 2008 National Football League season (*The Sports Network*).

(a) Fit a multiple regression model to relate the quarterback rating to the percentage of completions, the percentage of TDs, and the percentage of interceptions.
(b) Estimate σ^2.
(c) What are the standard errors of the regression coefficients?
(d) Use the model to predict the rating when the percentage of completions is 60%, the percentage of TDs is 4%, and the percentage of interceptions is 3%.

TABLE • E12-11 Quarterback Ratings for the 2008 National Football League Season

Player		Team	Att	Comp	Pct Comp	Yds	Yds per Att	TD	Pct TD	Lng	Int	Pct Int	Rating Pts
Philip	Rivers	SD	478	312	65.3	4,009	8.39	34	7.1	67	11	2.3	105.5
Chad	Pennington	MIA	476	321	67.4	3,653	7.67	19	4.0	80	7	1.5	97.4
Kurt	Warner	ARI	598	401	67.1	4,583	7.66	30	5.0	79	14	2.3	96.9
Drew	Brees	NO	635	413	65	5,069	7.98	34	5.4	84	17	2.7	96.2
Peyton	Manning	IND	555	371	66.8	4,002	7.21	27	4.9	75	12	2.2	95
Aaron	Rodgers	GB	536	341	63.6	4,038	7.53	28	5.2	71	13	2.4	93.8
Matt	Schaub	HOU	380	251	66.1	3,043	8.01	15	3.9	65	10	2.6	92.7
Tony	Romo	DAL	450	276	61.3	3,448	7.66	26	5.8	75	14	3.1	91.4
Jeff	Garcia	TB	376	244	64.9	2,712	7.21	12	3.2	71	6	1.6	90.2
Matt	Cassel	NE	516	327	63.4	3,693	7.16	21	4.1	76	11	2.1	89.4
Matt	Ryan	ATL	434	265	61.1	3,440	7.93	16	3.7	70	11	2.5	87.7
Shaun	Hill	SF	288	181	62.8	2,046	7.10	13	4.5	48	8	2.8	87.5
Seneca	Wallace	SEA	242	141	58.3	1,532	6.33	11	4.5	90	3	1.2	87
Eli	Manning	NYG	479	289	60.3	3,238	6.76	21	4.4	48	10	2.1	86.4
Donovan	McNabb	PHI	571	345	60.4	3,916	6.86	23	4.0	90	11	1.9	86.4
Jay	Cutler	DEN	616	384	62.3	4,526	7.35	25	4.1	93	18	2.9	86
Trent	Edwards	BUF	374	245	65.5	2,699	7.22	11	2.9	65	10	2.7	85.4
Jake	Delhomme	CAR	414	246	59.4	3,288	7.94	15	3.6	65	12	2.9	84.7
Jason	Campbell	WAS	506	315	62.3	3,245	6.41	13	2.6	67	6	1.2	84.3
David	Garrard	JAC	535	335	62.6	3,620	6.77	15	2.8	41	13	2.4	81.7
Brett	Favre	NYJ	522	343	65.7	3,472	6.65	22	4.2	56	22	4.2	81
Joe	Flacco	BAL	428	257	60	2,971	6.94	14	3.3	70	12	2.8	80.3
Kerry	Collins	TEN	415	242	58.3	2,676	6.45	12	2.9	56	7	1.7	80.2
Ben	Roethlisberger	PIT	469	281	59.9	3,301	7.04	17	3.6	65	15	3.2	80.1
Kyle	Orton	CHI	465	272	58.5	2,972	6.39	18	3.9	65	12	2.6	79.6
JaMarcus	Russell	OAK	368	198	53.8	2,423	6.58	13	3.5	84	8	2.2	77.1
Tyler	Thigpen	KC	420	230	54.8	2,608	6.21	18	4.3	75	12	2.9	76
Gus	Frerotte	MIN	301	178	59.1	2,157	7.17	12	4.0	99	15	5.0	73.7
Dan	Orlovsky	DET	255	143	56.1	1,616	6.34	8	3.1	96	8	3.1	72.6
Marc	Bulger	STL	440	251	57	2,720	6.18	11	2.5	80	13	3.0	71.4
Ryan	Fitzpatrick	CIN	372	221	59.4	1,905	5.12	8	2.2	79	9	2.4	70
Derek	Anderson	CLE	283	142	50.2	1,615	5.71	9	3.2	70	8	2.8	66.5

Att	Attempts (number of pass attempts)	Pct TD	Percentage of attempts that are touchdowns
Comp	Completed passes	Long	Longest pass completion
Pct Comp	Percentage of completed passes	Int	Number of interceptions
Yds	Yards gained passing	Pct Int	Percentage of attempts that are interceptions
Yds per Att	Yards gained per pass attempt	Rating Pts	Rating points
TD	Number of touchdown passes		

(a) The model can be expressed as:

```
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int
```

Regression Analysis: Rating Pts versus Pct Comp, Pct TD, Pct Int

```
The regression equation is
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int

Predictor      Coef   SE Coef      T      P
Constant      2.986     5.877   0.51  0.615
Pct Comp    1.19857   0.09743  12.30  0.000
Pct TD       4.5956    0.3848  11.94  0.000
```

```
Pct Int    -3.8125   0.4861  -7.84  0.000

S = 2.03479   R-Sq = 95.3%   R-Sq(adj) = 94.8%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  2373.59  791.20  191.09  0.000
Residual Error  28   115.93    4.14
Total           31  2489.52

Source    DF   Seq SS
Pct Comp   1  1614.43
Pct TD     1   504.49
Pct Int    1   254.67
```

(b) $\hat{\sigma}^2 = \dfrac{\sum_{t=1}^{n} e_t^2}{n-p} = \dfrac{SS_E}{n-p} = \dfrac{115.93}{28} = 4.14$

(c) $\text{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 C$, $se(\hat{\beta}) = \sqrt{\hat{\sigma}^2 C_{jj}} = \begin{bmatrix} 5.88 \\ 0.097 \\ 0.38 \\ 0.48 \end{bmatrix}$

from the SE Coef column in the computer output.

```
(d) Rating Pts = 2.99 + 1.20*60 + 4.60*4 - 3.81*3 = 81.96
```

12-22 **Regression Analysis: W versus GF, GA, ...**

Table E12-12 presents statistics for the National Hockey League teams from the 2008–2009 season (*The Sports Network*). Fit a multiple linear regression model that relates *wins* to the variables *GF* through *F*. Because teams play 82 game, $W = 82 - L - T - OTL$, but such a model does not help build a better team. Estimate σ^2 and find the standard errors of the regression coefficients for your model.

TABLE • E12-12 Team Statistics for the 2008–2009 National Hockey League Season

Team	W	L	OTL	PTS	GF	GA	ADV	PPGF	PCTG	PEN	BMI	AVG	SHT	PPGA	PKPCT	SHGF	SHGA	FG
Anaheim	42	33	7	91	238	235	309	73	23.6	1418	8	17.4	385	78	79.7	6	6	43
Atlanta	35	41	6	76	250	279	357	69	19.3	1244	12	15.3	366	88	76	13	9	39
Boston	53	19	10	116	270	190	313	74	23.6	1016	12	12.5	306	54	82.4	8	7	47
Buffalo	41	32	9	91	242	229	358	75	21	1105	16	13.7	336	61	81.8	7	4	44
Carolina	45	30	7	97	236	221	374	70	18.7	786	16	9.8	301	59	80.4	8	7	39
Columbus	41	31	10	92	220	223	322	41	12.7	1207	20	15	346	62	82.1	8	9	41
Calgary	46	30	6	98	251	246	358	61	17	1281	18	15.8	349	58	83.4	6	13	37
Chicago	46	24	12	104	260	209	363	70	19.3	1129	28	14.1	330	64	80.6	10	5	43
Colorado	32	45	5	69	190	253	318	50	15.7	1044	18	13	318	64	79.9	4	5	31
Dallas	36	35	11	83	224	251	351	54	15.4	1134	10	14	327	70	78.6	2	2	38
Detroit	51	21	10	112	289	240	353	90	25.5	810	14	10	327	71	78.3	6	4	46
Edmonton	38	35	9	85	228	244	354	60	17	1227	20	15.2	338	76	77.5	3	8	39
Florida	41	30	11	93	231	223	308	51	16.6	884	16	11	311	54	82.6	7	6	39
Los Angeles	34	37	11	79	202	226	360	69	19.2	1191	16	14.7	362	62	82.9	4	7	39
Minnesota	40	33	9	89	214	197	328	66	20.1	869	20	10.8	291	36	87.6	9	6	39
Montreal	41	30	11	93	242	240	374	72	19.2	1223	6	15	370	65	82.4	10	10	38
New Jersey	51	27	4	106	238	207	307	58	18.9	1038	20	12.9	324	65	79.9	12	3	44
Nashville	40	34	8	88	207	228	318	50	15.7	982	12	12.1	338	59	82.5	9	8	41
NY Islanders	26	47	9	61	198	274	320	54	16.9	1198	18	14.8	361	73	79.8	12	5	37
NY Rangers	43	30	9	95	200	212	346	48	13.9	1175	24	14.6	329	40	87.8	9	13	42
Ottawa	36	35	11	83	213	231	339	66	19.5	1084	14	13.4	346	64	81.5	8	5	46
Philadelphia	44	27	11	99	260	232	316	71	22.5	1408	26	17.5	393	67	83	16	1	43
Phoenix	36	39	7	79	205	249	344	50	14.5	1074	18	13.3	293	68	76.8	5	4	36
Pittsburgh	45	28	9	99	258	233	360	62	17.2	1106	8	13.6	347	60	82.7	7	11	46
San Jose	53	18	11	117	251	199	360	87	24.2	1037	16	12.8	306	51	83.3	12	10	46
St. Louis	41	31	10	92	227	227	351	72	20.5	1226	22	15.2	357	58	83.8	10	8	35
Tampa Bay	24	40	18	66	207	269	343	61	17.8	1280	26	15.9	405	89	78	4	8	34
Toronto	34	35	13	81	244	286	330	62	18.8	1113	12	13.7	308	78	74.7	6	7	40
Vancouver	45	27	10	100	243	213	357	67	18.8	1323	28	16.5	371	69	81.4	7	5	47
Washington	50	24	8	108	268	240	337	85	25.2	1021	20	12.7	387	75	80.6	7	9	45

W Wins
L Losses during regular time
OTL Overtime losses
PTS Points. Two points for winning a game, one point for a tie or losing in overtime, zero points for losing in regular time.
GF Goals for
GA Goals against
ADV Total advantages. Power-play opportunities.
PPGF Power-play goals for. Goals scored while on power play.
PCTG Power-play percentage. Power-play goals divided by total advantages.
PEN Total penalty minutes including bench minutes
BMI Total bench minor minutes
AVG Average penalty minutes per game
SHT Total times short-handed. Measures opponent opportunities.
PPGA Power-play goals against
PKPCT Penalty killing percentage. Measures a team's ability to prevent goals while its opponent is on a power play. Opponent opportunities minus power-play goals divided by opponent's opportunities.
SHGF Short-handed goals for
SHGA Short-handed goals against
FG Games scored first

```
The regression equation is
W = 512 + 0.164 GF - 0.183 GA - 0.054 ADV + 0.09 PPGF - 0.14 PCTG - 0.163 PEN
    - 0.128 BMI + 13.1 AVG + 0.292 SHT - 1.60 PPGA - 5.54 PKPCT + 0.106 SHGF
    + 0.612 SHGA + 0.005 FG

Predictor       Coef   SE Coef      T      P
Constant       512.2     185.9   2.75  0.015
GF           0.16374   0.03673   4.46  0.000
GA          -0.18329   0.04787  -3.83  0.002
ADV          -0.0540    0.2183  -0.25  0.808
PPGF           0.089     1.126   0.08  0.938
PCTG          -0.142     3.810  -0.04  0.971
PEN          -0.1632    0.3029  -0.54  0.598
BMI          -0.1282    0.2838  -0.45  0.658
AVG            13.09     24.84   0.53  0.606
SHT           0.2924    0.1334   2.19  0.045
```

```
PPGA         -1.6018   0.6407  -2.50  0.025
PKPCT         -5.542    2.181  -2.54  0.023
SHGF          0.1057   0.1975   0.54  0.600
SHGA          0.6124   0.2615   2.34  0.033
FG            0.0047   0.1943   0.02  0.981

S = 2.65443   R-Sq = 92.9%   R-Sq(adj) = 86.3%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression      14  1390.310  99.308  14.09  0.000
Residual Error  15   105.690   7.046
Total           29  1496.000
```

$$\hat{y} = 512.2 + 0.16374x_1 - 0.18329x_2 - 0.054x_3 + 0.089x_4 - 0.142x_5 - 0.1632x_6 - 0.1282x_7$$

$$+13.09x_8 + 0.2924x_9 - 1.6018x_{10} - 5.542x_{11} + 0.1057x_{12} + 0.6124x_{13} + 0.0047x_{14}$$

where

$$x_1 = GF \quad x_2 = GA \quad x_3 = ADV \quad x_4 = PPGF \quad x_5 = PCTG \quad x_6 = PEN \quad x_7 = BMI$$

$$x_8 = AVG \quad x_9 = SHT \quad x_{10} = PPGA \quad x_{11} = PKPCT \quad x_{12} = SHGF \quad x_{13} = SHGA \quad x_{14} = FG$$

$$\hat{\sigma}^2 = 7.046$$

The standard errors of the coefficients are listed under the SE Coef column above.

12-23 A study was performed on wear of a bearing and its relationship to x_1 = oil viscosity and x_2 = load. The following data were obtained.

y	x_1	x_2
293	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

(a) Fit a multiple linear regression model to these data.
(b) Estimate σ^2 and the standard errors of the regression coefficients.
(c) Use the model to predict wear when x_1 = 25 and x_2 = 1000.
(d) Fit a multiple linear regression model with an interaction term to these data.
(e) Estimate σ^2 and $se(\hat{\beta}_j)$ for this new model. How did these quantities change? Does this tell you anything about the value of adding the interaction term to the model?
(f) Use the model in part (d) to predict when x_1 = 25 and x_2 = 1000. Compare this prediction with the predicted value from part (c).

```
Predictor        Coef     SE Coef        T       P
Constant       383.80       36.22    10.60   0.002
X1            -3.6381      0.5665    -6.42   0.008
X2           -0.11168     0.04338    -2.57   0.082
S = 12.35       R-Sq = 98.5%    R-Sq(adj) = 97.5%

Analysis of Variance
Source           DF        SS        MS        F       P
Regression        2     29787     14894    97.59   0.002
Residual Error    3       458       153
Total             5     30245
```

(a) $\hat{y} = 383.80 - 3.6381x_1 - 0.1117x_2$

(b) $\hat{\sigma}^2 = 153.0$, $se(\hat{\beta}_0) = 36.22$, $se(\hat{\beta}_1) = 0.5665$, and $se(\hat{\beta}_2) = .04338$

(c) $\hat{y} = 383.80 - 3.6381(25) - 0.1119(1000) = 180.95$

(d)

```
Predictor      Coef     SE Coef          T        P
Constant        484.0      101.3       4.78    0.041
X1             -7.656      3.846      -1.99    0.185
X2            -0.2221     0.1129      -1.97    0.188
X1*X2        0.004087   0.003871       1.06    0.402
S = 12.12       R-Sq = 99.0%     R-Sq(adj) = 97.6%
Analysis of Variance
Source            DF          SS          MS         F        P
Regression         3     29951.4      9983.8     67.92    0.015
Residual Error     2       294.0       147.0
Total              5     30245.3
```

$$\hat{y} = 484.0 - 7.656x_1 - 0.222x_2 - 0.0041x_{12}$$

(e) $\hat{\sigma}^2 = 147.0$, $se(\hat{\beta}_0) = 101.3$, $se(\hat{\beta}_1) = 3.846$, $se(\hat{\beta}_2) = 0.113$ and $se(\hat{\beta}_{12}) = 0.0039$

(f) $\hat{y} = 484.0 - 7.656(25) - 0.222(1000) - 0.0041(25)(1000) = 173.1$

The predicted value is smaller.

12-24 Consider the linear regression model

$$Y_i = \beta'_0 + \beta'(x_{i1} - \overline{x_1}) + \beta_2(x_{i2} - \overline{x_2}) + \in_i$$

where $\overline{x_1} = \sum x_{i1}/n$ and $\overline{x_2} = \sum x_{i2}/n$.

(a) Write out the least squares normal equations for this model.

(b) Verify that the least squares estimate of the intercept in this model is $\hat{\beta}'_0 = \sum y_i / n = \overline{y}$.

(c) Suppose that we use $y_i - \overline{y}$ as the response variable in this model. What effect will this have on the least squares estimate of the intercept?

(a) $f(\beta'_0, \beta_1, \beta_2) = \sum [y_i - \beta'_0 - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)]^2$

$$\frac{\partial f}{\partial \beta'_0} = -2\sum [y_i - \beta'_0 - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)]$$

$$\frac{\partial f}{\partial \beta_1} = -2\sum [y_i - \beta'_0 - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)](x_{i1} - \bar{x}_1)$$

$$\frac{\partial f}{\partial \beta_2} = -2\sum [y_i - \beta'_0 - \beta_1(x_{i1} - \bar{x}_1) - \beta_2(x_{i2} - \bar{x}_2)](x_{i2} - \bar{x}_2)$$

Setting the derivatives equal to zero yields

$$n\beta_0^{'} = \sum y_i$$
$$n\beta_0^{'} + \beta_1 \sum (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = \sum y_i (x_{i1} - \bar{x}_1)$$
$$n\beta_0^{'} + \beta_1 \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) + \beta_2 \sum (x_{i2} - \bar{x}_2)^2 = \sum y_i (x_{i2} - \bar{x}_2)$$

(b) From the first normal equation, $\hat{\beta}_0^{'} = \bar{y}$.

(c) Substituting $y_i - \bar{y}$ for y_i in the first normal equation yields $\hat{\beta}_0^{'} = 0$.

Sections 12-2

12-25 Recall the regression of percent of body fat on height and waist from Exercise 12-1. The simple regression model of percent of body fat on height alone shows the following:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	25.58078	14.15400	1.807	0.0719
Height	-0.09316	0.20119	-0.463	0.6438

(a) Test whether the coefficient of height is statistically significant.
(b) Looking at the model with both waist and height in the model, test whether the coefficient of height is significant in this model.
(c) Explain the discrepancy in your two answers.

(a) Because the *P*-value = 0.6438 > 0.05, the coefficient that corresponds to *height* is not statistically significant.

(b) From the computer output in the referenced exercise, the *p*-value = 1.09e-07 < 0.05. Therefore the coefficient that corresponds to *height* is statistically significant.

(c) A simple regression model with *height* alone as a predictor variable does not detect that the variable is statistically significant. However, *height* contributes significantly to a model that has both *height* and *waist* as predictor variables.

12-26 Exercise 12-2 presented a regression model to predict final grade from two hourly tests.

(a) Test the hypotheses that each of the slopes is zero.
(b) What is the value of R^2 for this model?
(c) What is the residual standard deviation?
(d) Do you believe that the professor can predict the final grade well enough from the two hourly tests to consider not giving the final exam? Explain.

A value is needed for $y'y$. Assume $y'y = \sum_{i=1}^{63} y_i^2 = 411222.7041$
Then SSE = $y'y - y'X(X'X)^{-1}X'y$ = 411222.7041 – 403874.0241 = 7348.68
An estimate of σ^2 is 7348.68/60 = 122.478
$(X'X)^{-1}$ =

0.912917	-9.82E-03	-7.11E-04
-0.00982	1.50E-04	-4.16E-05
-7.11E-04	-4.16E-05	5.81E-05

$X'y$ =

4871
426011
367576.5

Covariance matrix of $\hat{\beta} = \sigma^2 (X'X)^{-1}$
Var $\hat{\beta}_1$=122.478(1.50E-04) = 0.0184
Var $\hat{\beta}_2$=122.478(5.81E-05) = 0.00712

$t = \frac{\hat{\beta}_1}{SE\hat{\beta}_1} = \frac{0.691136}{\sqrt{0.0184}} = 5.10$

$$t = \frac{\hat{\beta}_{12}}{SE\hat{\beta}_2} = \frac{0..186642}{\sqrt{0.00712}} = 2.21$$

n = 63, so degrees of freedom = 63 – 3 = 60

For β_1 the P-value < 0.001, reject H_0 at $\alpha = 0.01$
For β_2 the P-value = 0.031, reject H_0 at $\alpha = 0.05$

(b) R^2 = 1 – SSE/SST = 1 - 7348.68/403996.5021 = 0.982

(c) Residual standard deviation = $122.478^{1/2}$ = 11.067

(d) Yes, R^2 is large

12-27 Consider the regression model of Exercise 12-3 attempting to predict the percent of engineers in the workforce from various spending variables.

(a) Are any of the variables useful for prediction? (Test an appropriate hypothesis).
(b) What percent of the variation in the percent of engineers is accounted for by the model?
(c) What might you do next to create a better model?

(a) $$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_1: \beta_j \neq 0 \quad \text{for at least one } j$$

The F-statistic is 2.968 with *P*-value = 0.04157 < .0.05. Therefore we can reject the null hypothesis at $\alpha = 0.05$, and conclude that at least one of the variables is significant.

(b) From the output of the regression model, we see the multiple R-squared = 0.1622, so 16.22% of the variation in the percentage of engineers is accounted for by the model.

(c) We might consider the original variables raised to a certain power (such as a second order model) or cross-product terms between the original variables. Furthermore, additional variables might be useful to improve the model.

12-28 Consider the linear regression model from Exercise 12-4. Is the second-order term necessary in the regression model?

Computer output for the model with *ratio* and *ratio* squared is shown below. The *P*-value for the test of whether the coefficient of *ratio squared* equals zero is 0.309 > 0.05. Therefore, there is not sufficient evidence that the *ratio squared* variable is useful to the model.

```
The regression equation is
viscosity = 0.198 + 1.37 ratio - 1.28 ratio2

Predictor    Coef  SE Coef      T      P
Constant   0.1979   0.4466   0.44  0.676
ratio       1.367    1.488   0.92  0.400
ratio2     -1.280    1.131  -1.13  0.309

S = 0.146606   R-Sq = 37.5%   R-Sq(adj) = 12.5%

Analysis of Variance

Source          DF       SS       MS     F      P
Regression       2  0.06442  0.03221  1.50  0.309
Residual Error   5  0.10747  0.02149
Total            7  0.17189
```

12-29 Consider the following computer output.

The regression equation is Y = 254 + 2.77 x1 – 3.58 x2

Predictor	Coef	SE Coef	T	P
Constant	253.810	4.781	?	?
x1	2.7738	0.1846	15.02	?
x2	–3.5753	0.1526	?	?

S = 5.05756 R-Sq = ? R-Sq (adj) = 98.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	22784	11392	?	?
Residual error	?	?	?		
Total	14	23091			

(a) Fill in the missing quantities. You may use bounds for the *P*-values.
(b) What conclusions can you draw about the significance of regression?
(c) What conclusions can you draw about the contributions of the individual regressors to the model?

(a) $t_0 = \dfrac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$, null hypothesis $\hat{\beta}_j = \beta_{j0}$ is rejected at α level if $|t_0| > t_{\alpha/2, n-p}$

$F_0 = \dfrac{SS_R / k}{SS_E /(n-p)} = \dfrac{MS_R}{MS_E}$, regression is significant at α level if $f_0 > f_{\alpha,k,n-p}$

The missing quantities are as follows:

Predictor	Coef	SE Coef	T	P	
Constant	253.81	4.781	53.0872	0	
x1	2.7738	0.1846	15.02	0	
x2	-3.5753	0.1526	-23.4292	0	
Source	DF	SS	MS	F	P
Regression	2	22784	11392	445.2899	0
Residual Error	12	307	25.5833		
Total	14	23091			

R-Squared = 22784/23091 = 0.9867

(b) From the P-value from the F test (F = 445.2899) for regression is significant.

(c) Each individual regressor is significant to the model that contains the other regressors.

12-30 You have fit a regression model with two regressors to a data set that has 20 observations. The total sum of squares is 1000 and the model sum of squares is 750.

(a) What is the value of R^2 for this model?
(b) What is the adjusted R^2 for this model?
(c) What is the value of the *F*-statistic for testing the significance of regression? What conclusions would you draw about this model if α = 0.05? What if α = 0.01?
(d) Suppose that you add a third regressor to the model and as a result, the model sum of squares is now 785. Does it seem to you that adding this factor has improved the model?

(a) $R^2 = \dfrac{SS_R}{SS_T} = \dfrac{750}{1000} = 0.75$

(b) $SS_E = SS_T - SS_R = 1000 - 750 = 250$

$$R^2_{adj} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \frac{250/(20-3)}{1000/(20-1)} = 0.7206$$

(c) $MS_{Regression} = \dfrac{SS_{Regression}}{k} = \dfrac{750}{2} = 375$

$MS_{Error} = \dfrac{SS_E}{n-p} = \dfrac{1000-750}{17} = \dfrac{250}{17} = 14.7059$

$F = \dfrac{MS_{Regression}}{MS_{Error}} = \dfrac{375}{14.7059} = 25.5$

The ANOVA table

Source	DF	SS	MS	F	P
Regression	2	750	375	25.5	< 0.01
Residual Error	17	250	14.7059		
Total	19	1000			

For the F test the P-value < 0.0. Therefore the F test rejects the null hypothesis at $\alpha = 0.05$ and also rejects at $\alpha = 0.01$.

(d) The ANOVA table after adding a third regressor

Source	DF	SS	MS
Regression	3	785	261.6667
Residual Error	16	215	13.44
Total	19	1000	

$f = \dfrac{SS_{Regression}(\beta_3 \mid \beta_2, \beta_1, \beta_0)/1}{MS_{Error}} = \dfrac{785-750}{13.44} = 2.60$

Because $f_{0.05,1,16} = 4.49$, we fail to reject H_0 and conclude that the third regressor does not contribute significantly to the model.

12-31 Consider the regression model fit to the soil shear strength data in Exercise 12-5.

(a) Test for significance of regression using $\alpha = 0.05$. What is the *P*-value for this test?
(b) Construct the *t*-test on each regression coefficient. What are your conclusions, using $\alpha = 0.05$? Calculate *P*-values.

(a) n = 10, k = 2, p = 3, $\alpha = 0.05$

$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$

$H_1: \beta_j \neq 0$ for at least one j

The calculations are sensitive to round off and the following calculations use more digits than are shown in the intermediate steps to obtain the final answers.

$$S_{yy} = 371595.6 - \frac{(1916)^2}{10} = 4490$$

$$X'y = \begin{bmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \sum x_{i2} y_i \end{bmatrix} = \begin{bmatrix} 1916 \\ 43550.8 \\ 104736.8 \end{bmatrix}$$

$$\hat{\beta}' X' y = \begin{bmatrix} 171.055 & 3.713 & -1.126 \end{bmatrix} \begin{bmatrix} 1916 \\ 43550.8 \\ 104736.8 \end{bmatrix} = 371536.689$$

$$SS_R = 371536.689 - \frac{1916^2}{10} = 4431.1$$

$$SS_E = S_{yy} - SS_R = 4490 - 4431.1 = 58.9$$

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{4431.1/2}{58.9/7} = 263.26$$

$$f_{0.05,2,7} = 4.74$$

$$f_0 > f_{0.05,2,7}$$

Reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$. P-value ≈ 0.000

(b) $\hat{\sigma}^2 = MS_E = \dfrac{SS_E}{n-p} = \dfrac{58.9}{7} = 8.416$

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{11}} = \sqrt{8.416(0.00439)} = 0.192$$

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{22}} = \sqrt{8.416(0.00087)} = 0.086$$

$H_0 : \beta_1 = 0$	$\beta_2 = 0$
$H_1 : \beta_1 \neq 0$	$\beta_2 \neq 0$
$t_0 = \dfrac{\hat{\beta}_1}{se(\hat{\beta}_1)}$	$t_0 = \dfrac{\hat{\beta}_2}{se(\hat{\beta}_2)}$
$= \dfrac{3.713}{0.192} = 19.32$	$= \dfrac{-1.126}{0.10199} = -13.16$
$t_{\alpha/2,7} = t_{0.025,7} = 2.365$	
Reject H_0, P-value < 0.001	Reject H_0, P-value < 0.001

Both regression coefficients significant

12-32 Consider the absorption index data in Exercise 12-6. The total sum of squares for y is $SS_T = 742.00$.

(a) Test for significance of regression using $\alpha = 0.01$. What is the P-value for this test?
(b) Test the hypothesis $H_0: \beta_1 = 0$ versus $H_1: \beta_1 \neq 0$ using $\alpha = 0.01$. What is the P-value for this test?
(c) What conclusion can you draw about the usefulness of $x1$ as a regressor in this model?

$$S_{yy} = 742.00$$

(a) $H_0 : \beta_1 = \beta_2 = 0$

$H_1 : \beta_j \neq 0$ for at least one j

$\alpha = 0.01$

$$SS_R = \hat{\beta}' X' y - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$= \begin{pmatrix} -1.9122 & 0.0931 & 0.2532 \end{pmatrix} \begin{pmatrix} 220 \\ 36768 \\ 9965 \end{pmatrix} - \frac{220^2}{10}$$

$$= 5525.5548 - 4840$$

$$= 685.55$$

$$SS_E = S_{yy} - SS_R$$

$$= 742 - 685.55$$

$$= 56.45$$

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{685.55/2}{56.45/7} = 42.51$$

$$f_{0.01,2,7} = 9.55$$

$$f_0 > f_{0.01,2,7}$$

Reject H_0 and conclude that the regression model is significant at $\alpha = 0.01$. P-value = 0.000121

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-p} = \frac{56.45}{7} = 8.0643$$

$$se(\hat{\beta}_1) = \sqrt{8.0643(7.9799E-5)} = 0.0254$$

(b) $H_0 : \beta_1 = 0$

$$H_1 : \beta_1 \neq 0$$

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

$$= \frac{0.0931}{0.0254} = 3.67$$

$$t_{0.005,7} = 3.499$$

$$|t_0| > t_{0.005,7}$$

Reject H_0 and conclude that β_1 is significant in the model at $\alpha = 0.01$

P-value = $2(1 - P(t < t_0))$ = 2(1 - 0.996018) = 0.007964

(c) x_1 is useful as a regressor in the model.

12-33 A regression model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$ as been fit to a sample of n = 25 observations. The calculated t-ratios $\hat{\beta}_j / se(\hat{\beta}_j)$, j =1, 2, 3 are as follows: for β_1, $t_0 = 4.82$, for β_2, $t_0 = 8.21$, and for β_3, $t_0 = 0.98$.

(a) Find P-values for each of the t-statistics.
(b) Using α= 0.05, what conclusions can you draw about the regressor x_3? Does it seem likely that this regressor contributes significantly to the model?

(a) Degrees of freedom = 25 – 4 = 21

$\beta_1 : t_0 = 4.82$ P-value = 2(4.589 E-5) = 9.18 E-5

$\beta_2 : t_0 = 8.21$ P-value = 2(2.711 E-8) = 5.42 E-8

$\beta_3 : t_0 = 0.98$ P-value = 2 (0.1691) = 0.338

(b) $H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

$\alpha = 0.05$

$t_0 = 0.98$, P-value = 2 (0.1691) = 0.338

Because the P-value > $\alpha = 0.05$, fail to reject H_0. We conclude that x_3 does not contribute significantly to the model.

12-34 Consider the electric power consumption data in Exercise 12-10.

(a) Test for significance of regression using $\alpha = 0.05$. What is the *P*-value for this test?
(b) Use the *t*-test to assess the contribution of each regressor to the model. Using $\alpha = 0.05$, what conclusions can you draw?

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_1 at least one $\beta_j \neq 0$

$\alpha = 0.05$

$f_0 = 10.08$

$f_{0.05,4,7} = 4.12$

$f_0 > f_{0.05,4,7}$

Reject H_0 P-value = 0.005

(b) $\alpha = 0.05$

$H_0 : \beta_1 = 0$ $\beta_2 = 0$ $\beta_3 = 0$ $\beta_4 = 0$

$H_1 : \beta_1 \neq 0$ $\beta_2 \neq 0$ $\beta_3 \neq 0$ $\beta_4 \neq 0$

$t_0 = 2.71$ $t_0 = 1.87$ $t_0 = 1.37$ $t_0 = -0.87$

$t_{\alpha/2,n-p} = t_{0.025,7} = 2.365$

$|t_0| \not> t_{0.0257}$ for β_2, β_3 and β_4

Reject H_0 for β_1.

12-35 Consider the gasoline mileage data in Exercise 12-11.

(a) Test for significance of regression using $\alpha = 0.05$. What conclusions can you draw?
(b) Find the *t*-test statistic for each regressor. Using $\alpha = 0.05$, what conclusions can you draw? Does each regressor contribute to the model?

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

H_1: at least one $\beta \neq 0$

$f_0 = 19.53$

$f_{\alpha,6,14} = f_{0.05,6,14} = 2.848$

$f_0 > f_{0.05,6,14}$

Reject H_0 and conclude regression model is significant at $\alpha = 0.05$

(b) The t-test statistics for β_1 through β_6 are -0.45, -0.07, -3.42, 0.17, -2.90, 0.69. Because $t_{0.025,14} = 2.14$, the regressors that contribute to the model at $\alpha = 0.05$ are *etw* and *axle*.

12-36 Consider the wire bond pull strength data in Exercise 12-12.

(a) Test for significance of regression using $\alpha = 0.05$. Find the *P*-value for this test. What conclusions can you draw?
(b) Calculate the *t*-test statistic for each regression coefficient. Using $\alpha = 0.05$, what conclusions can you draw? Do all variables contribute to the model?

(a) $H_0 : \beta_j = 0$ for all j

$H_1 : \beta_j \neq 0$ for at least one j

$f_0 = 7.16$

$f_{.05,4,14} = 3.11$

$f_0 > f_{0.05,4,14}$

Reject H_0 and conclude that the regression is significant at $\alpha = 0.05$. P-value = 0.0023

(b) $\hat{\sigma} = 0.7792$

$\alpha = 0.05$ $t_{\alpha/2,n-p} = t_{.025,14} = 2.145$

$H_0{:}\beta_2 = 0$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = 0$
$H_1{:}\beta_2 \neq 0$	$\beta_3 \neq 0$	$\beta_4 \neq 0$	$\beta_5 \neq 0$
$t_0 = -0.113$	$t_0 = 3.83$	$t_0 = -1.91$	$t_0 = -0.9$
$\lvert t_0\rvert \not> t_{\alpha/2,14}$	$\lvert t_0\rvert > t_{\alpha/2,14}$	$\lvert t_0\rvert \not> t_{\alpha/2,14}$	$\lvert t_0\rvert \not> t_{\alpha/2,14}$
Fail to reject H_0	Reject H_0	Fail to reject H_0	Fail to reject H_0

All variables do not contribute to the model.

12-37 Reconsider the semiconductor data in Exercise 12-13.

(a) Test for significance of regression using $\alpha = 0.05$. What conclusions can you draw?
(b) Calculate the *t*-test statistic and *P*-value for each regression coefficient. Using $\alpha = 0.05$, what conclusions can you draw?

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$H_1{:}\beta_j \neq 0$ for at least one j

$f_0 = 828.31$

$f_{.05,3,16} = 3.34$

$f_0 > f_{0.05,3,16}$

Reject H_0 and conclude regression is significant at $\alpha = 0.05$

(b) $\hat{\sigma}^2 = 12.2856$

$\alpha = 0.05$ $t_{\alpha/2,n-p} = t_{.025,16} = 2.12$

$H_0{:}\beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$
$H_1{:}\beta_1 \neq 0$	$\beta_2 \neq 0$	$\beta_3 \neq 0$
$t_0 = -2.58$	$t_0 = 1.84$	$t_0 = 13.82$
$\lvert t_0 \rvert > t_{0.025,16}$	$\lvert t_0 \rvert \not> t_{0.025,16}$	$\lvert t_0 \rvert > t_{0.025,16}$
Reject H_0	Fail to reject H_0	Reject H_0

12-38 Consider the regression model fit to the arsenic data in Exercise 12-16. Use arsenic in nails as the response and age, drink use, and cook use as the regressors.

(a) Test for significance of regression using α = 0.05. What is the *P*-value for this test?
(b) Construct a *t*-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use α = 0.05.

```
ARSNAILS = 0.001 + 0.00858 AGE - 0.021 DRINKUSE + 0.010 COOKUSE

Predictor       Coef   SE Coef      T      P
Constant      0.0011    0.9067   0.00  0.999
AGE         0.008581  0.007083   1.21  0.242
DRINKUSE     -0.0208    0.1018  -0.20  0.841
COOKUSE       0.0097    0.1798   0.05  0.958

S = 0.506197   R-Sq = 8.1%   R-Sq(adj) = 0.0%

Analysis of Variance

Source          DF      SS      MS     F      P
Regression       3  0.3843  0.1281  0.50  0.687
Residual Error  17  4.3560  0.2562
Total           20  4.7403
```

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
$H_1 : \beta_j \neq 0$
for at least one j; k=4

$\alpha = 0.05$
$f_0 = 0.50$
$f_{0.05,3,17} = 3.197$
$f_0 < f_{0.05,3,17}$

Fail to reject H_0. There is insufficient evidence to conclude that the model is significant at α = 0.05. The *P*-value = 0.687.

(b) $H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$
α = 0.05

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{0.008581}{0.007083} = 1.21$$

$t_{0.025,17} = 2.11$

$|t_0| < t_{\alpha/2,17}$. Fail to reject H_0, there is not enough evidence to conclude that β_1 is significant in the model at α = 0.05.

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$
α = 0.05

$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{-0.0208}{0.1018} = -0.2$$

$t_{0.025,17} = 2.11$

$|t_0| < t_{\alpha/2,17}$. Fail to reject H_0, there is not enough evidence to conclude that β_2 is significant in the model at α = 0.05.

$H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

α = 0.05

$$t_0 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{0.0097}{0.1798} = 0.05$$

$t_{0.025,17} = 2.11$

$|t_0| < t_{\alpha/2,17}$. Fail to reject H_0, there is not enough evidence to conclude that β_3 is significant in the model at α = 0.05.

12-39 Consider the regression model fit to the X-ray inspection data in Exercise 12-15. Use rads as the response.

(a) Test for significance of regression using α = 0.05. What is the *P*-value for this test?
(b) Construct a *t*-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use α = 0.05.

(a) $H_0 : \beta_1 = \beta_2 = 0$

H_0 :for at least one $\beta_j \neq 0$

α = 0.05

$f_0 = 99.67$

$f_{0.05,2,37} = 3.252$

$f_0 > f_{0.05,2,37}$

```
 The regression equation is
rads = - 440 + 19.1 mAmps + 68.1 exposure time


Predictor          Coef  SE Coef      T      P
Constant        -440.39    94.20  -4.68  0.000
mAmps            19.147    3.460   5.53  0.000
exposure time    68.080    5.241  12.99  0.000


S = 235.718   R-Sq = 84.3%   R-Sq(adj) = 83.5%


Analysis of Variance

Source          DF        SS       MS      F      P
Regression       2  11076473  5538237  99.67  0.000
Residual Error  37   2055837    55563
Total           39  13132310
```

Reject H_0 and conclude regression model is significant at α = 0.05. P-value < 0.000001

(b) $\hat{\sigma}^2 = MS_E = 55563$

$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 3.460$

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$\alpha = 0.05$

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
$$= \frac{19.147}{3.460} = 5.539$$
$$t_{0.025,40-3} = t_{0.025,37} = 2.0262$$
$$|t_0| > t_{\alpha/2,37},$$

Reject H_0 and conclude that β_1 is significant in the model at $\alpha = 0.05$

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 5.241$$
$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$
$\alpha = 0.05$

$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$
$$= \frac{68.080}{5.241} = 12.99$$
$$t_{0.025,40-3} = t_{0.025,37} = 2.0262$$
$$|t_0| > t_{\alpha/2,37},$$

Reject H_0 conclude that β_2 is significant in the model at $\alpha = 0.05$

12-40 Consider the regression model fit to the nisin extraction data in Exercise 12-18. Use nisin extraction as the response.

(a) Test for significance of regression using $\alpha = 0.05$. What is the *P*-value for this test?
(b) Construct a *t*-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use $\alpha = 0.05$.
(c) Comment on the effect of a small sample size to the tests in the previous parts.

```
The regression equation is
y = - 171 + 7.03 x1 + 12.7 x2

Predictor      Coef  SE Coef      T      P
Constant    -171.26    28.40  -6.03  0.001
x1            7.029    1.539   4.57  0.004
x2           12.696    1.539   8.25  0.000


S = 3.07827   R-Sq = 93.7%   R-Sq(adj) = 91.6%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  842.37  421.18  44.45  0.000
Residual Error   6   56.85    9.48
Total            8  899.22
```

(a) $H_0: \beta_1 = \beta_2 = 0$

$H_1:$ for at least one $\beta_j \neq 0$

α = 0.05

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{842.37/2}{56.85/6} = 44.45$$

$$f_{0.05,2,6} = 5.14$$

$$f_0 > f_{0.05,2,6}$$

Reject H_0 and conclude regression model is significant at α = 0.05 P-value ≈ 0

(b) $\hat{\sigma}^2 = MS_E = 9.48$

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 1.539$$

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

α = 0.05

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

$$= \frac{7.03}{1.539} = 4.568$$

$$t_{0.025,9-3} = t_{0.025,6} = 2.447$$

$$|t_0| > t_{\alpha/2,6},$$

Reject H_0, β_1 is significant in the model at α = 0.05

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 1.539$$

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

α = 0.05

$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$

$$= \frac{12.7}{1.539} = 8.252$$

$$t_{0.025,9-3} = t_{0.025,6} = 2.447$$

$$|t_0| > t_{\alpha/2,6},$$

Reject H_0 conclude that β_2 is significant in the model at α = 0.05

(c) With a smaller sample size, the difference in the estimate from the hypothesized value needs to be greater to be significant.

12-41 Consider the regression model fit to the gray range modulation data in Exercise 12-19. Use the useful range as the response.

(a) Test for significance of regression using $\alpha = 0.05$. What is the *P*-value for this test?
(b) Construct a *t*-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use $\alpha = 0.05$.

```
Useful range (ng) = 239 + 0.334 Brightness (%) - 2.72 Contrast (%)

Predictor          Coef  SE Coef      T      P
Constant         238.56    45.23   5.27  0.002
Brightness (%)   0.3339   0.6763   0.49  0.639
Contrast (%)    -2.7167   0.6887  -3.94  0.008

S = 36.3493   R-Sq = 75.6%   R-Sq(adj) = 67.4%

Analysis of Variance

Source          DF     SS     MS     F      P
Regression       2  24518  12259  9.28  0.015
Residual Error   6   7928   1321
Total            8  32446
```

(a) $H_0 : \beta_1 = \beta_2 = 0$

$H_1 :$ for at least one $\beta_j \neq 0$

$\alpha = 0.05$

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{24518/2}{7928/6} = 9.28$$

$$f_{0.05,2,6} = 5.14$$

$$f_0 > f_{0.05,2,6}$$

Reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$ P-value =0.015

(b) $\hat{\sigma}^2 = MS_E = 1321$

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.6763$$

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

$\alpha = 0.05$

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

$$= \frac{0.3339}{0.6763} = 0.49$$

$$t_{0.025,9-3} = t_{0.025,6} = 2.447$$

$$|t_0| < t_{\alpha/2,6},$$

Fail to reject H_0, there is no enough evidence to conclude that β_1 is significant in the model at $\alpha = 0.05$

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.6887$$
$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$
$\alpha = 0.05$
$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$
$$= \frac{-2.7167}{0.6887} = -3.94$$
$$t_{0.025,9-3} = t_{0.025,6} = 2.447$$
$$|t_0| > t_{\alpha/2,6},$$
Reject H_0 conclude that β_2 is significant in the model at $\alpha = 0.05$

12-42 Consider the regression model fit to the stack loss data in Exercise 12-20. Use stack loss as the response.

(a) Test for significance of regression using $\alpha = 0.05$. What is the *P*-value for this test?
(b) Construct a *t*-test on each regression coefficient. What conclusions can you draw about the variables in this model? Use $\alpha = 0.05$.

```
The regression equation is
Stack Loss(y) = - 39.9 + 0.716 X1 + 1.30 X2 - 0.152 X3

Predictor     Coef  SE Coef      T      P
Constant    -39.92    11.90  -3.36  0.004
X1          0.7156   0.1349   5.31  0.000
X2          1.2953   0.3680   3.52  0.003
X3         -0.1521   0.1563  -0.97  0.344

S = 3.24336   R-Sq = 91.4%   R-Sq(adj) = 89.8%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3  1890.41  630.14  59.90  0.000
Residual Error  17   178.83   10.52
Total           20  2069.24
```

(a) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$H_1 : \beta_j \neq 0$ for at least one j

$\alpha = 0.05$

$$f_0 = \frac{\frac{SS_R}{k}}{\frac{SS_E}{n-p}} = \frac{189.41/3}{178.83/17} = 59.90$$
$$f_{0.05,3,17} = 3.20$$
$$f_0 > f_{0.05,3,17}$$
Reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$ P-value < 0.000001

(b) $\hat{\sigma}^2 = MS_E = 10.52$

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.1349$$
$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
α = 0.05
$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
$$= \frac{0.7156}{0.1349} = 5.31$$
$$t_{0.025,21-4} = t_{0.025,17} = 2.110$$
$$|t_0| > t_{\alpha/2,17}.$$
Reject H_0 and conclude that β_1 is significant in the model at α = 0.05.

$$se(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.3680$$
$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$
α = 0.05
$$t_0 = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$
$$= \frac{1.2953}{0.3680} = 3.52$$
$$t_{0.025,21-4} = t_{0.025,17} = 2.110$$
$|t_0| > t_{\alpha/2,17}$. Reject H_0 and conclude that β_2 is significant in the model at α = 0.05.

$$se(\hat{\beta}_3) = \sqrt{\hat{\sigma}^2 c_{jj}} = 0.1563$$
$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$
α = 0.05
$$t_0 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)}$$
$$= \frac{-0.1521}{0.1563} = -0.97$$
$$t_{0.025,21-4} = t_{0.025,17} = 2.110$$
$$|t_0| < t_{\alpha/2,17}.$$
Fail to reject H_0, there is not enough evidence to conclude that β_3 is significant in the model at α = 0.05.

12-43 Consider the NFL data in Exercise 12-21.

(a) Test for significance of regression using α = 0.05. What is the *P*-value for this test?
(b) Construct a *t*-test on each regression coefficient. What conclusions can you draw about the variables in this model?

Use α = 0.05.

(c) Find the amount by which the regressor x_2 (TD percentage) increases the regression sum of squares, and conduct an *F*-test for $H_0{:}\,\beta_2 = 0$ versus $H_1{:}\,\beta_2 \neq 0$ using α = 0.05. What is the *P*-value for this test? What conclusions can you draw?

(a) Computer output follows. The test statistic is F = 191.09. Because the P-value is near zero, the regression is significant at $\alpha = 0.05$.

(b) $t_0 = \dfrac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$, null hypothesis $\hat{\beta}_j = \beta_{j0}$ is rejected at α level if $|t_0| > t_{\alpha/2,n-p}$ or the P-value < α

The P-values of all regressors are less than 0.05. Therefore, all individual variables in the model are significant.

(c) The computer output for three regressors is followed by the computer output for two regressors. From the regression sum of squares in each model the F test for x_2 is

$$F_0 = \frac{SS_R(\beta_1 \mid \beta_2)/r}{MS_E} = \frac{2373.59 - 1782.96}{4.14} = 142.66$$

The F-test P-value is near zero. Therefore the regressor (TD percentage) is significant to the model. This is the equivalent to the *t* test on the coefficient of x_2. The F statistic = 142.66 = 11.94^2, except for some round-off error.

Results of regression on three variables and on two variables are shown below.

Regression Analysis: Rating Pts versus Pct Comp, Pct TD, Pct Int

```
The regression equation is
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int

Predictor      Coef  SE Coef      T      P
Constant      2.986    5.877   0.51  0.615
Pct Comp    1.19857  0.09743  12.30  0.000
Pct TD       4.5956   0.3848  11.94  0.000
Pct Int     -3.8125   0.4861  -7.84  0.000

S = 2.03479   R-Sq = 95.3%   R-Sq(adj) = 94.8%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  2373.59  791.20  191.09  0.000
Residual Error  28   115.93    4.14
Total           31  2489.52

Source    DF   Seq SS
Pct Comp   1  1614.43
Pct TD     1   504.49
Pct Int    1   254.67


Unusual Observations

Obs  Pct Comp  Rating Pts     Fit  SE Fit  Residual  St Resid
 11      61.1      87.700  83.691   0.371     4.009     2.00R
 18      59.4      84.700  79.668   0.430     5.032     2.53R
 21      65.7      81.000  85.020   1.028    -4.020    -2.29R
 31      59.4      70.000  75.141   0.719    -5.141    -2.70R
```

```
R denotes an observation with a large standardized residual.
```

Regression Analysis: Rating Pts versus Pct Comp, Pct Int

```
The regression equation is
Rating Pts = - 9.1 + 1.66 Pct Comp - 3.08 Pct Int

Predictor    Coef  SE Coef      T      P
Constant    -9.11    14.04  -0.65  0.522
Pct Comp   1.6622   0.2168   7.67  0.000
Pct Int    -3.076    1.170  -2.63  0.014

S = 4.93600   R-Sq = 71.6%   R-Sq(adj) = 69.7%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2  1782.96  891.48  36.59  0.000
Residual Error  29   706.56   24.36
Total           31  2489.52
```

12-44 Exercise 12-14 presents data on heat-treating gears.

(a) Test the regression model for significance of regression. Using α = 0.05, find the *P*-value for the test and draw conclusions.
(b) Evaluate the contribution of each regressor to the model using the *t*-test with α = 0.05.
(c) Fit a new model to the response PITCH using new regressors x_1 = SOAKTIME × SOAKPCT and x_2 = DIFFTIME × DIFFPCT.
(d) Test the model in part (c) for significance of regression using α = 0.05. Also calculate the *t*-test for each regressor and draw conclusions.
(e) Estimate σ^2 for the model from part (c) and compare this to the estimate of σ^2 for the model in part (a). Which estimate is smaller? Does this offer any insight regarding which model might be preferable?

(a) $H_0: \beta_j = 0$ for all j
$H_1: \beta_j \neq 0$ for at least one j

$f_0 = 158.9902$

$f_{.05,5,26} = 2.59$

$f_0 > f_{\alpha,5,26}$

Reject H_0 and conclude regression is significant at α = 0.05.

P-value < 0.000001

(b) α = 0.05 $t_{\alpha/2,n-p} = t_{.025,26} = 2.056$

$H_0: \beta_1 = 0$	$\beta_2 = 0$	$\beta_3 = 0$	$\beta_4 = 0$	$\beta_5 = 0$
$H_1: \beta_1 \neq 0$	$\beta_2 \neq 0$	$\beta_3 \neq 0$	$\beta_4 \neq 0$	$\beta_5 \neq 0$
$t_0 = 0.83$	$t_0 = 12.25$	$t_0 = -0.52$	$t_0 = 6.96$	$t_0 = -0.29$
Fail to reject H_0	Reject H_0	Fail to reject H_0	Reject H_0	Fail to reject H_0

(c) $\hat{y} = 0.010889 + 0.002687x_1 + 0.009325x_2$

(d) $H_0: \beta_j = 0$ for all j
$H_1: \beta_j \neq 0$ for at least one j

$f_0 = 308.455$

$f_{.05,2,29} = 3.33$

$f_0 > f_{0.05,2,29}$

Reject H_0 and conclude regression is significant at $\alpha = 0.05$

$\alpha = 0.05$	$t_{\alpha/2,n-p} = t_{.025,29} = 2.045$
$H_0 : \beta_1 = 0$	$\beta_2 = 0$
$H_1 : \beta_1 \neq 0$	$\beta_2 \neq 0$
$t_0 = 18.31$	$t_0 = 6.37$
$\lvert t_0 \rvert > t_{\alpha/2,29}$	$\lvert t_0 \rvert > t_{\alpha/2,29}$

Reject H_0 for each regressor variable and conclude that both variables are significant at $\alpha = 0.05$

(e) $\hat{\sigma}_{part(d)} = 6.7E-6$.

Part c) is smaller, suggesting a better model.

12-45 Consider the bearing wear data in Exercise 12-23.

(a) For the model with no interaction, test for significance of regression using $\alpha = 0.05$. What is the *P*-value for this test? What are your conclusions?
(b) For the model with no interaction, compute the *t*-statistics for each regression coefficient. Using $\alpha = 0.05$, what conclusion can you draw?
(c) For the model with no interaction, use the extra sum of squares method to investigate the usefulness of adding x_2 = load to a model that already contains x_1 = oil viscosity. Use $\alpha = 0.05$.
(d) Refit the model with an interaction term. Test for significance of regression using $\alpha = 0.05$.
(e) Use the extra sum of squares method to determine whether the interaction term contributes significantly to the model. Use $\alpha = 0.05$.
(f) Estimate σ^2 for the interaction model. Compare this to the estimate of σ^2 from the model in part (a).

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, Assume no interaction model.

(a) $H_0 : \beta_1 = \beta_2 = 0$

H_1 at least one $\beta_j \neq 0$

$f_0 = 97.59$

$f_{0.05,2,3} = 9.55$

$f_0 > f_{0.05,2,3}$

Reject H_0. P-value = 0.002

(b) $H_0 : \beta_1 = 0$	$H_0 : \beta_2 = 0$
$H_1 : \beta_1 \neq 0$	$H_1 : \beta_2 \neq 0$
$t_0 = -6.42$	$t_0 = -2.57$
$t_{\alpha/2,3} = t_{0.025,3} = 3.182$	$t_{\alpha/2,3} = t_{0.025,3} = 3.182$
$\lvert t_0 \rvert > t_{0.025,3}$	$\lvert t_0 \rvert \not> t_{0.025,3}$
Reject H_0 for regressor β_1.	Fail to reject H_0 for regressor β_2.

(c) $SS_R(\beta_2 \mid \beta_1, \beta_0) = 1012$

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$
α = 0.05
$f_0 = 6.629$
$f_{\alpha,1,3} = f_{.05,1,3} = 10.13$
$f_0 \ngtr f_{0.05,1,3}$
Fail to reject H_0

(d) $H_0 : \beta_1 = \beta_2 = \beta_{12} = 0$
H1 at least one $\beta_j \neq 0$
α = 0.05
$f_0 = 7.714$
$f_{\alpha,3,2} = f_{0.05,3,2} = 19.16$
$f_0 \ngtr f_{0.05,3,2}$
Fail to reject H_0

(e) $H_0 : \beta_{12} = 0$
$H_1 : \beta_{12} \neq 0$
α = 0.05
$SSR(\beta_{12} \mid \beta_1, \beta_2) = 29951.4 - 29787 = 163.9$
$f_0 = \dfrac{SSR}{MS_E} = \dfrac{163.9}{147} = 1.11$
$f_{0.05,1,2} = 18.51$
$f_0 \ngtr f_{0.05,1,2}$
Fail to reject H_0

(f) $\hat{\sigma}^2 = 147.0$
$\hat{\sigma}^2$ (no interaction term) = 153.0
$MS_E(\hat{\sigma}^2)$ was reduced in the model with the interaction term.

12-46 Data on National Hockey League team performance were presented in Exercise 12-22.

(a) Test the model from this exercise for significance of regression using α = 0.05. What conclusions can you draw?
(b) Use the *t*-test to evaluate the contribution of each regressor to the model. Does it seem that all regressors are necessary? Use α = 0.05.
(c) Fit a regression model relating the number of games won to the number of goals for and the number of power play goals for. Does this seem to be a logical choice of regressors, considering your answer to part (b)? Test this new model for significance of regression and evaluate the contribution of each regressor to the model using the *t*-test. Use α = 0.05.

(a) $H_0 : \beta_j = 0$ for all j
$H_1 : \beta_j \neq 0$ for at least one j
From the computer output
$f_0 = 14.09$
$f_{0.05,14,15} = 2.424$
$f_0 > f_{0.05,14,15}$
Reject H_0 and conclude that the regression model is significant at α = 0.05

(b) $H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

$t_{.025,15} = 2.131$

GF :	$t_0 = 4.46$	Reject H_0
GA :	$t_0 = -3.83$	Reject H_0
ADV :	$t_0 = -0.25$	Fail to reject H_0
PPGF :	$t_0 = 0.08$	Fail to reject H_0
PCTG :	$t_0 = -0.04$	Fail to reject H_0
PEN :	$t_0 = -0.54$	Fail to reject H_0
BMI :	$t_0 = -0.45$	Fail to reject H_0
AVG :	$t_0 = 0.53$	Fail to reject H_0
SHT :	$t_0 = 2.19$	Reject H_0
PPGA :	$t_0 = -2.50$	Reject H_0
PKPCT :	$t_0 = -2.54$	Reject H_0
SHGF :	$t_0 = 0.54$	Fail to reject H_0
SHGA :	$t_0 = 2.34$	Reject H_0
FG :	$t_0 = 0.02$	Fail to reject H_0

It does not seem that all regressors are important. Only the regressors "GF" (β_1), "GA" (β_2), "SHT" (β_9), "PPGA" (β_{10}), "PKPCT" (β_{11}), and "SHGA" (β_{13}) are significant at $\alpha = 0.05$

(c) The computer result is shown below.

Regression Analysis: W versus GF, PPGF

```
The regression equation is
W = - 8.82 + 0.218 GF - 0.016 PPGF

Predictor      Coef  SE Coef      T      P
Constant     -8.818    9.230  -0.96  0.348
GF          0.21779  0.05467   3.98  0.000
PPGF        -0.0162   0.1134  -0.14  0.888

S = 5.11355   R-Sq = 52.8%   R-Sq(adj) = 49.3%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2   789.99  395.00  15.11  0.000
Residual Error  27   706.01   26.15
Total           29  1496.00
```

Because PPGF had a *t* statistic near zero in part (b) there is a concern that it is not an important predictor. We will evaluate its role in the smaller model with GF.

$$\hat{y} = -8.82 + 0.218x_1 - 0.16x_4$$

$$f_0 = 15.11$$

$$f_{0.05,2,27} = 3.35$$

Because $f_0 > f_{0.05,2,27}$, we reject the null hypothesis that the coefficient of GF and PPGF are both zero.

$H_0 : \beta_1 = 0$	$\beta_4 = 0$
$H_1 : \beta_1 \neq 0$	$\beta_4 \neq 0$
$t_0 = 3.98$	$t_0 = -0.14$
Reject H_0	Fail to reject H_0

Based on the t-test, power play goals for (PPGF) is not a logical choice to add to the model that already contains GF.

12-47 Data from a hospital patient satisfaction survey were presented in Exercise 12-9.

(a) Test the model from this exercise for significance of regression. What conclusions can you draw if $\alpha = 0.05$? What if $\alpha = 0.01$?
(b) Test the contribution of the individual regressors using the *t*-test. Does it seem that all regressors used in the model are really necessary?

(a) The computer output follows. The P-value for the F-test is near zero. Therefore, the regression is significant at both $\alpha = 0.05$ or $\alpha = 0.01$

(b) $t_0 = \dfrac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)}$. Because the P-values for Age and Severity are < 0.05 both regressors are significant to the model. Because the P-value for Anxiety is 0.233, it is not significant to the model at level $\alpha = 0.05$.

Regression Analysis: Satisfaction versus Age, Severity, Anxiety

```
The regression equation is
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety

Predictor      Coef  SE Coef      T      P
Constant    143.895    5.898  24.40  0.000
Age         -1.1135   0.1326  -8.40  0.000
Severity    -0.5849   0.1320  -4.43  0.000
Anxiety       1.296    1.056   1.23  0.233

S = 7.03710   R-Sq = 90.4%   R-Sq(adj) = 89.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3   9738.3  3246.1  65.55  0.000
Residual Error  21   1039.9    49.5
Total           24  10778.2

Source    DF  Seq SS
Age        1  8756.7
Severity   1   907.0
Anxiety    1    74.6
```

12-48 Data from a hospital patient satisfaction survey were presented in Exercise 12-9.

(a) Fit a regression model using only the patient age and severity regressors. Test the model from this exercise for significance of regression. What conclusions can you draw if $\alpha = 0.05$? What if $\alpha = 0.01$?
(b) Test the contribution of the individual regressors using the *t*-test. Does it seem that all regressors used in the model are really necessary?
(c) Find an estimate of the error variance σ^2. Compare this estimate of σ^2 with the estimate obtained from the model

containing the third regressor, anxiety. Which estimate is smaller? Does this tell you anything about which model might be preferred?

(a) The computer output is shown below.

Regression Analysis: Satisfaction versus Age, Severity

```
The regression equation is
Satisfaction = 143 - 1.03 Age - 0.556 Severity

Predictor     Coef  SE Coef      T      P
Constant   143.472    5.955  24.09  0.000
Age        -1.0311   0.1156  -8.92  0.000
Severity   -0.5560   0.1314  -4.23  0.000

S = 7.11767   R-Sq = 89.7%   R-Sq(adj) = 88.7%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2   9663.7  4831.8  95.38  0.000
Residual Error  22   1114.5    50.7
Total           24  10778.2
```

Because the P-value of the F test is less than $\alpha = 0.05$ and $\alpha = 0.01$, we reject the H_0 and conclude that at least one regressor contributes significantly to the model at either α level.

(b) Because the P-values from the t-test for both *age* and *severity* regressors are less than $\alpha = 0.05$, we reject the H_0 and conclude that both *age* and *severity* regressors contribute significantly to the model.

(c) From $MS_{Residual}$, the estimate of the variance = 50.7. From the computer output below, if the third variable *anxiety* is added to the model, the estimate of the variance is reduced to 49.5. The variance changed very slightly here so it is unlikely that the variable contributes significantly to the model.

Regression Analysis: Satisfaction versus Age, Severity, Anxiety

```
The regression equation is
Satisfaction = 144 - 1.11 Age - 0.585 Severity + 1.30 Anxiety

Predictor     Coef  SE Coef      T      P
Constant   143.895    5.898  24.40  0.000
Age        -1.1135   0.1326  -8.40  0.000
Severity   -0.5849   0.1320  -4.43  0.000
Anxiety      1.296    1.056   1.23  0.233

S = 7.03710   R-Sq = 90.4%   R-Sq(adj) = 89.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3   9738.3  3246.1  65.55  0.000
Residual Error  21   1039.9    49.5
Total           24  10778.2
```

<u>Sections 12-3 and 12-4</u>

12-49 Using the regression model from Exercise 12-1,

(a) Find a 95% confidence interval for the coefficient of height.
(b) Find a 95% confidence interval for the mean percent of body fat for a man with a height of 72in and waist of 34in.
(c) Find a 95% prediction interval for the percent of body fat for a man with the same height and waist as in part (b).
(d) Which interval is wider, the confidence interval or the prediction interval? Explain briefly.
(e) Given your answer to part (c), do you believe that this is a useful model for predicting body fat? Explain briefly.

$$t_{\alpha/2,247} = 1.97, \text{ for } \alpha = 0.05$$

(a)

$$\hat{\beta}_1 - t \times SE\hat{\beta}_1 \le \beta_1 \le \hat{\beta}_1 + t \times SE\hat{\beta}_1$$
$$-0.60154 - 1.97 \times 0.10994 \le \beta_1 \le -0.60154 + 1.97 \times 0.10994$$
$$-0.818 \le \beta_1 \le -0.385$$

(b) $x_0 = (1, 72, 34)$, $x_0'(X'X)^{-1}x_0 = 0.007866$ and $\hat{\sigma} = 4.46$ from the computer output

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_{21}$$
$$\hat{y} = -3.10088 - 0.60154(72) + 1.77309(34) = 13.8733$$

$$13.8733 - 1.97 \times \sqrt{4.46^2(0.007866)} \le \mu_{Y|x_0} \le 13.8733 + 1.97 \times \sqrt{4.46^2(0.007866)}$$
$$13.094 \le \mu_{Y|x_0} \le 14.653$$

(c) $13.8733 - 1.97 \times \sqrt{4.46^2(1 + 0.007866)} \le \hat{Y} \le 13.8733 + 1.97 \times \sqrt{4.46^2(1 + 0.007866)}$

$$5.053 \le \hat{Y} \le 22.694$$

(d) The prediction interval is wider. The confidence interval expresses the uncertainty in estimating the mean of a distribution while prediction interval expresses the uncertainty in predicting a future observation from a distribution.

(e) The prediction interval is quite wide. Therefore, the model is not useful to predict body fat.

12-50 Using the regression from Exercise 12-2,

(a) Find a 95% confidence interval for the coefficient of hourly 1 test.
(b) Find a 95% confidence interval for the mean final grade for students who score 80 on the first test and 85 on the second.
(c) Find a 95% prediction interval for a student with the same grades as in part (b).

(a) $t_{0.05/2,60} = 2.00, \text{ for } \alpha = 0.05$

$$\hat{\beta}_1 - t \times SE\hat{\beta}_1 \le \beta_1 \le \hat{\beta}_1 + t \times SE\hat{\beta}_1$$
$$0.691 - 2.00 \times \sqrt{0.0184} \le \beta_1 \le 0.691 + 2.00 \times \sqrt{0.0184}$$
$$0.420 \le \beta_1 \le 0.962$$

(b) $x_0 = (1, 80, 85)$, $x_0'(X'X)^{-1}x_0 = 0.0349$ and $\hat{\sigma} = 122.478^{1/2} = 11.067$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_{21}$$
$$\hat{y} = 4.076 - 0.691(80) + 0.187(85) = 75.231$$

$$75.231 - 2.00 \times \sqrt{122.478(0.0349)} \le \mu_{Y|x_0} \le 75.231 + 2.00 \times \sqrt{122.478(0.0349)}$$
$$71.098 \le \mu_{Y|x_0} \le 79.365$$

(c)

$$75.231 - 2.00 \times \sqrt{122.478(1+0.0349)} \le \mu_{Y|x_0} \le 75.231 + 2.00 \times \sqrt{122.478(1+0.0349)}$$

$52.712 \le \hat{Y} \le 97.751$

12-51 Referring to the regression model from Exercise 12-3,

(a) Find a 95% confidence interval for the coefficient of spending on higher education.
(b) Is zero in the confidence interval you found in part (a)? What does that fact imply about the coefficient of higher education?
(c) Find a 95% prediction interval for a state that has \$1 per \$1000 in venture capital, spends \$10,000 per student on funding for major research universities, and spends 0.5% of its GDP on higher education.

$$t_{\alpha/2,46} = 2.013, \text{for } \alpha = 0.05$$

(a) $-0.1673 - 2.013 \times 0.2595 \le \beta_3 \le -0.1673 + 2.013 \times 0.2595$

$$-0.690 \le \beta_3 \le 0.355$$

(b) Yes. *Higher Ed* is not a significant variable.

(c) Refers to a prediction interval that cannot be computed from the given info.

12-52 Use the second-order polynomial regression model from Exercise 12-4,

(a) Find a 95% confidence interval on both the first-order and the second-order term in this model.
(b) Is zero in the confidence interval you found for the second-order term in part (a)? What does that fact tell you about the contribution of the second-order term to the model?
(c) Refit the model with only the first-order term. Find a 95% confidence interval on this term. Is this interval longer or shorter than the confidence interval that you found on this term in part (a)?

Computer output for the second-order model follows.

```
The regression equation is
viscosity = 0.198 + 1.37 ratio - 1.28 ratio2

Predictor    Coef  SE Coef      T      P
Constant   0.1979   0.4466   0.44  0.676
ratio       1.367    1.488   0.92  0.400
ratio2     -1.280    1.131  -1.13  0.309

S = 0.146606   R-Sq = 37.5%   R-Sq(adj) = 12.5%

Analysis of Variance

Source          DF       SS        MS     F      P
Regression       2  0.06442  0.03221  1.50  0.309
Residual Error   5  0.10747  0.02149
Total            7  0.17189
```

(a)

$$t_{\alpha/2,467} = 2.365 \text{ for } \alpha = 0.05$$

$$1.367 - 2.365 \times 1.488 \le \beta_1 \le 1.367 + 2.365 \times 1.488$$

$$-2.152 \le \beta_1 \le 4.886$$

$$-1.280 - 2.365 \times 1.131 \le \beta_2 \le -1.280 + 2.365 \times 1.131$$

$$-3.955 \le \beta_2 \le 1.395$$

(b) Yes. The second-order term is not a significant variable.

(c) Computer output for the first-order model follows.

```
The regression equation is
viscosity = 0.671 - 0.296 ratio

Predictor     Coef  SE Coef      T      P
Constant    0.6714   0.1595   4.21  0.006
ratio      -0.2964   0.2314  -1.28  0.248

S = 0.149990   R-Sq = 21.5%   R-Sq(adj) = 8.4%

Analysis of Variance

Source          DF       SS       MS     F      P
Regression       1  0.03691  0.03691  1.64  0.248
Residual Error   6  0.13498  0.02250
Total            7  0.17189
```

$$-0.2964-2.306\times 0.2314\le \beta_1\le -0.2964+2.306\times 0.2314$$
$$-3.498\le \beta_1\le -2.430$$

This confidence interval is much narrower. The standard error of the estimated coefficient is much lower in this model. This illustrates that a first-order model that uses the significant variables can improve these estimates.

12-53 Consider the regression model fit to the shear strength of soil in Exercise 12-5.

(a) Calculate 95% confidence intervals on each regression coefficient.
(b) Calculate a 95% confidence interval on mean strength when x_1 = 18 ft and x_2 = 43%.
(c) Calculate 95% prediction interval on strength for the same values of the regressors used in the previous part.

(a) $\hat{\beta}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{00}}$

$$171.055 \pm t_{.025,7}\ se(\hat{\beta}_0)$$
$$171.055 \pm (2.365)(51.217)$$
$$171.055 \pm 121.128$$
$$49.927 \le \beta_0 \le 292.183$$

$$\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$$
$$3.713 \pm t_{.025,7}\ se(\hat{\beta}_1)$$
$$3.713 \pm (2.365)(1.556)$$
$$3.713 \pm 3.680$$
$$0.033 \le \beta_1 \le 7.393$$

$$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$$
$$-1.126 \pm t_{.025,7}\ se(\hat{\beta}_0)$$
$$-1.126 \pm (2.365)(0.693)$$
$$-1.126 \pm 1.639$$
$$-2.765 \le \beta_2 \le 0.513$$

(b) $x_1 = 18$

$x_2 = 43$

$\hat{y}_0 = 189.471$

$X_0^{'}(X'X)^{-1}X_0 = 0.305065$

$189.471 \pm (2.365)\sqrt{550.7875(0.305065)}$

$158.815 \le \mu_{Y|x_0} \le 220.127$

(c) α = 0.05

$x_1 = 18$

$x_2 = 43$

$\hat{y}_0 = 189.471$

$X_0^{'}(X'X)^{-1}X_0 = 0.305065$

$\hat{y} \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1 + X_0^{'}(X'X)^{-1}X_0)}$

$189.471 \pm (2.365)\sqrt{550.7875(1.305065)}$

$126.064 \le y_0 \le 252.878$

12-54 Consider the soil absorption data in Exercise 12-6.

(a) Find 95% confidence intervals on the regression coefficients.
(b) Find a 95% confidence interval on mean soil absorption index when x_1 = 200 and x_2 = 50.
(c) Find a 95% prediction interval on the soil absorption index when x_1 = 200 and x_2 = 50.

(a) $\hat{\beta}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{00}}$

$-1.9122 \pm t_{.025,7}\ se(\hat{\beta}_0)$

$-1.9122 \pm (2.365)(10.055)$

-1.9122 ± 23.78

$-25.6922 \le \beta_0 \le 21.8678$

$\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$

$0.0931 \pm t_{.025,7}\ se(\hat{\beta}_1)$

$0.0931 \pm (2.365)(0.0827)$

0.0931 ± 0.1956

$-0.1025 \le \beta_1 \le 0.2887$

$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$

$0.2532 \pm t_{.025,7}\ se(\hat{\beta}_0)$

$0.2532 \pm (2.365)(0.1998)$

0.2532 ± 0.4725

$-0.2193 \le \beta_2 \le 0.7257$

(b) $\mathrm{x}_1 = 200$

$x_2 = 50$

$\hat{y}_0 = 29.37$

$$X_0'(X'X)^{-1}X_0 = 0.211088$$
$$29.37 \pm (2.365)\sqrt{85.694(0.211088)}$$
$$29.37 \pm 10.059$$
$$19.311 \le \mu_{Y|x_0} \le 39.429$$

(c) $\alpha = 0.05$

$$x_1 = 200$$
$$x_2 = 50$$
$$\hat{y}_0 = 29.37$$
$$X_0'(X'X)^{-1}X_0 = 0.211088$$
$$\hat{y} \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+X_0'(X'X)^{-1}X_0)}$$
$$29.37 \pm (2.365)\sqrt{85.694(1.211088)}$$
$$29.37 \pm 24.093$$
$$5.277 \le y_0 \le 53.463$$

12-55 Consider the semiconductor data in Exercise 12-13.

(a) Find 99% confidence intervals on the regression coefficients.
(b) Find a 99% prediction interval on HFE when $x_1 = 14.5$, $x_2 = 220$, and $x_3 = 5.0$
(c) Find a 99% confidence interval on mean HFE when $x_1 = 14.5$, $x_2 = 220$, and $x_3 = 5.0$.

(a) $-20.477 \le \beta_1 \le 1.269$
$$-0.245 \le \beta_2 \le 1.076$$
$$14.428 \le \beta_3 \le 22.159$$

(b) $\hat{\mu}_{Y|x_0} = 91.372$ $\quad se(\hat{y}_0) = 4.721 \; t_{.005,16} = 2.921$
$$91.372 \pm 2.921(4.721)$$
$$77.582 \le y_0 \le 105.162$$

(c) $\hat{\mu}_{Y|x_0} = 91.372$ $\quad se(\hat{\mu}_{Y|x_0}) = 3.163$
$$91.372 \pm (2.921)(3.163)$$
$$82.133 \le \mu_{Y|x_0} \le 100.611$$

12-56 Consider the electric power consumption data in Exercise 12-10.

(a) Find 95% confidence intervals on β_1, β_2, β_3 and β_4.
(b) Find a 95% confidence interval on the mean of Y when $x_1 = 75$, $x_2 = 24$, $x_3 = 90$, and $x_4 = 98$.
(c) Find a 95% prediction interval on the power consumption when $x_1 = 75$, $x_2 = 24$, $x_3 = 90$, and $x_4 = 98$.

(a) 95 % CI on coefficients

$$0.0973 \le \beta_1 \le 1.4172$$
$$-1.9646 \le \beta_2 \le 17.0026$$
$$-1.7953 \le \beta_3 \le 6.7613$$
$$-1.7941 \le \beta_4 \le 0.8319$$

(b) $\hat{\mu}_{Y|x_0} = 290.44$ $se(\hat{\mu}_{Y|x_0}) = 7.61$ $t_{.025,7} = 2.365$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$$
$$290.44 \pm (2.365)(7.61)$$
$$272.44 \le \mu_{Y|x_0} \le 308.44$$

(c) $\hat{y}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1+\mathbf{x}_0{}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}$

$$290.44 \pm 2.365(14.038)$$
$$257.25 \le y_0 \le 323.64$$

12-57 Consider the bearing wear data in Exercise 12-23.

(a) Find 99% confidence intervals on β_1 and β_2.
(b) Recompute the confidence intervals in part (a) after the interaction term x_1x_2 is added to the model. Compare the lengths of these confidence intervals with those computed in part (a) . Do the lengths of these intervals provide any information about the contribution of the interaction term in the model?

(a) $-6.9467 \le \beta_1 \le -0.3295$

$$-0.3651 \le \beta_2 \le 0.1417$$

(b) $-45.8276 \le \beta_1 \le 30.5156$

$$-1.3426 \le \beta_2 \le 0.8984$$
$$-0.03433 \le \beta_{12} \le 0.04251$$

These part b) intervals are much wider.
Yes, the addition of this term increased the standard error of the regression coefficient estimators.

12-58 Consider the wire bond pull strength data in Exercise 12-12.

(a) Find 95% confidence interval on the regression coefficients.
(b) Find a 95% confidence interval on mean pull strength when $x_2 = 20$, $x_3 = 30$, $x_4 = 90$, and $x_5 = 2.0$.
(c) Find a 95% prediction interval on pull strength when $x_2 = 20$, $x_3 = 30$, $x_4 = 90$, and $x_5 = 2.0$.

(a) $-0.595 \le \beta_2 \le 0.535$

$$0.229 \le \beta_3 \le 0.812$$
$$-0.216 \le \beta_4 \le 0.013$$
$$-7.2968 \le \beta_5 \le 2.9756$$

(b) $\hat{\mu}_{Y|x_0} = 8.99568$ $se(\hat{\mu}_{Y|x_0}) = 0.472445$ $t_{.025,14} = 2.145$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$$
$$8.99568 \pm (2.145)(0.472445)$$
$$7.982 \le \mu_{Y|x_0} \le 10.009$$

(c) $\hat{y}_0 = 8.99568$ $se(\hat{y}_0) = 100121$

$$8.99568 \pm 2.145(1.00121)$$
$$6.8481 \le y_0 \le 11.143$$

12-59 Consider the regression model fit to the X-ray inspection data in Exercise 12-15. Use rads as the response.

(a) Calculate 95% confidence intervals on each regression coefficient.
(b) Calculate a 99% confidence interval on mean rads at 15 milliamps and 1 second on exposure time.
(c) Calculate a 99% prediction interval on rads for the same values of the regressors used in the part (b).

(a)

$$\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$$
$$19.147 \pm t_{.025,37}\ se(\hat{\beta}_1)$$
$$19.147 \pm (2.0262)(3.460)$$
$$19.147 \pm 7.014458$$
$$12.1363 \le \beta_1 \le 26.1577$$

$$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$$
$$68.080 \pm t_{.025,37}\ se(\hat{\beta}_2)$$
$$68.080 \pm (2.0262)(5.241)$$
$$68.080 \pm 7.014458$$
$$57.4607 \le \beta_2 \le 78.6993$$

(b)

$$\hat{\mu}_{Y|x_0} = -85.1 \qquad se(\hat{\mu}_{Y|x_0}) = 54.6 \qquad t_{0.005,37} = 2.7154$$
$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p}\ se(\hat{\mu}_{Y|x_0})$$
$$-85.1 \pm (2.7154)(54.6)$$
$$-233.4 \le \mu_{Y|x_0} \le 63.2$$

(c) $\hat{y}_0 = -85.1$ $\qquad se(\hat{y}_0) = 241.95$

$$-85.1 \pm 2.7154(241.95)$$
$$-742.09 \le y_0 \le 571.89$$

12-60 Consider the regression model fit to the arsenic data in Exercise 12-16. Use arsenic in nails as the response and age, drink use, and cook use as the regressors.

(a) Calculate 99% confidence intervals on each regression coefficient.
(b) Calculate a 99% confidence interval on mean arsenic concentration in nails when age = 30, drink use = 4, and cook Use = 4
(c) Calculate a prediction interval on arsenic concentration in nails for the same values of the regressors used in part (b).

```
The regression equation is
ARSNAILS = 0.001 + 0.00858 AGE - 0.021 DRINKUSE + 0.010 COOKUSE

Predictor       Coef    SE Coef      T      P
Constant      0.0011     0.9067   0.00  0.999
AGE         0.008581   0.007083   1.21  0.242
DRINKUSE     -0.0208     0.1018  -0.20  0.841
COOKUSE       0.0097     0.1798   0.05  0.958

S = 0.506197   R-Sq = 8.1%   R-Sq(adj) = 0.0%

Analysis of Variance
```

```
Source           DF      SS      MS     F       P
Regression        3  0.3843  0.1281  0.50  0.687
Residual Error   17  4.3560  0.2562
Total            20  4.7403
```

(a) $t_{0.005,17} = 2.898$

$-2.617 \le \beta_0 \le 2.634$

$-0.012 \le \beta_1 \le 0.0291$

$-0.316 \le \beta_2 \le 0.274$

$-0.511 \le \beta_3 \le 0.531$

(b) $\hat{\mu}_{Y|x_0} = 0.214$ $se(\hat{\mu}_{Y|x_0}) = 0.216$ $t_{0.005,16} = 2.898$

$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$

$0.214 \pm (2.898)(0.216)$

$-0.412 \le \mu_{Y|x_0} \le 0.840$

(c) $\hat{y}_0 = 0.214$ $se(\hat{y}_0) = 0.5503$

$0.214 \pm 2.898(0.5503)$

$-1.381 \le y_0 \le 1.809$

12-61 Consider the regression model fit to the coal and limestone mixture data in Exercise 12-17. Use density as the response.

(a) Calculate 90% confidence intervals on each regression coefficient.
(b) Calculate a 90% confidence interval on mean density when the dielectric constant = 2.3 and the loss factor = 0.025.
(c) Calculate a prediction interval on density for the same values of the regressors used in part (b).

(a) $t_{0.05,8} = 1.860$

$-0.576 \le \beta_0 \le 0.355$

$0.0943 \le \beta_1 \le 0.7201$

$-8.743 \le \beta_2 \le 12.959$

(b) $\hat{\mu}_{Y|x_0} = 0.8787$ $se(\hat{\mu}_{Y|x_0}) = 0.00926$ $t_{0.005,16} = 1.860$

$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$

$0.8787 \pm (1.860)(0.00926)$

$0.86148 \le \mu_{Y|x_0} \le 0.89592$

(c) $\hat{y}_0 = 0.8787$ $se(\hat{y}_0) = 0.0134$

$0.8787 \pm 1.86(0.0134)$

$0.85490 \le y_0 \le 0.90250$

12-62 Consider the regression model fit to the nisin extraction data in Exercise 12-18.

(a) Calculate 95% confidence intervals on each regression coefficient.
(b) Calculate a 95% confidence interval on mean nisin extraction when $x_1 = 15.5$ and $x_2 = 16$.

(c) Calculate a prediction interval on nisin extraction for the same values of the regressors used in part (b).
(d) Comment on the effect of a small sample size to the widths of these intervals.

```
The regression equation is
y = - 171 + 7.03 x1 + 12.7 x2

Predictor     Coef  SE Coef      T      P
Constant   -171.26    28.40  -6.03  0.001
x1           7.029    1.539   4.57  0.004
x2          12.696    1.539   8.25  0.000

S = 3.07827   R-Sq = 93.7%   R-Sq(adj) = 91.6%

Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  842.37  421.18  44.45  0.000
Residual Error   6   56.85    9.48
Total            8  899.22
```

(a) $\hat{\beta}_1 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{11}}$

$$7.03 \pm t_{.025,6}\ se(\hat{\beta}_1)$$
$$7.03 \pm (2.447)(1.539)$$
$$7.03 \pm 3.766$$
$$3.264 \le \beta_1 \le 10.796$$

$$\hat{\beta}_2 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 c_{22}}$$
$$12.7 \pm t_{.025,6}\ se(\hat{\beta}_2)$$
$$12.7 \pm (2.447)(1.539)$$
$$12.7 \pm 3.766$$
$$8.934 \le \beta_1 \le 16.466$$

(b)

```
New
Obs     Fit  SE Fit       95% CI             95% PI
  1  140.82    6.65  (124.54, 157.11)  (122.88, 158.77)XX
```

$$\hat{\mu}_{Y|x_0} = 140.82 \qquad se(\hat{\mu}_{Y|x_0}) = 6.65 \qquad t_{0.025,6} = 2.447$$
$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p}\ se(\hat{\mu}_{Y|x_0})$$
$$140.82 \pm (2.447)(6.65)$$
$$124.55 \le \mu_{Y|x_0} \le 157.09$$

(c) $\hat{y}_0 = 140.82$ $\qquad se(\hat{y}_0) = 7.33$

$$140.82 \pm 2.447(7.33)$$
$$122.88 \le y_0 \le 158.76$$

(d) The smaller the sample size, the wider the interval

12-63 Consider the regression model fit to the gray range modulation data in Exercise 12-19. Use the useful range as the response.

(a) Calculate 99% confidence intervals on each regression coefficient.
(b) Calculate a 99% confidence interval on mean useful range when brightness = 70 and contrast = 80.
(c) Calculate a prediction interval on useful range for the same values of the regressors used in part (b).
(d) Calculate a 99% confidence interval and a 99% a prediction interval on useful range when brightness = 50 and contrast = 25. Compare the widths of these intervals to those calculated in parts (b) and (c). Explain any differences in widths.

```
The regression equation is
Useful range (ng) = 239 + 0.334 Brightness (%) - 2.72 Contrast (%)

Predictor          Coef  SE Coef      T      P
Constant         238.56    45.23   5.27  0.002
Brightness (%)   0.3339   0.6763   0.49  0.639
Contrast (%)    -2.7167   0.6887  -3.94  0.008

S = 36.3493   R-Sq = 75.6%   R-Sq(adj) = 67.4%

Analysis of Variance

Source          DF     SS     MS     F      P
Regression       2  24518  12259  9.28  0.015
Residual Error   6   7928   1321
Total            8  32446
```

(a) $t_{0.005,6} = 3.707$

$$-2.173 \le \beta_1 \le 2.841$$

$$-5.270 \le \beta_2 \le -0.164$$

(b)

```
Predicted Values for New Observations

New
Obs   Fit  SE Fit      99% CI           99% PI
  1  44.6    21.9  (-36.7, 125.8)  (-112.8, 202.0)

Values of Predictors for New Observations

New                   Contrast
Obs  Brightness (%)        (%)
  1            70.0       80.0
```

$\hat{\mu}_{Y|x_0} = 44.6$ $\quad se(\hat{\mu}_{Y|x_0}) = 21.9$ $\quad t_{0.005,6} = 3.707$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \quad se(\hat{\mu}_{Y|x_0})$$

$$44.6 \pm (3.707)(21.9)$$

$$-36.7 \le \mu_{Y|x_0} \le 125.8$$

(c) $\hat{y}_0 = 44.6$ $\quad se(\hat{y}_0) = 42.44$

$$44.6 \pm 3.707(42.44)$$

$$-112.8 \le y_0 \le 202.0$$

(d) `Predicted Values for New Observations`

```
New
Obs    Fit  SE Fit       99% CI           99% PI
  1  187.3    21.6  (107.4, 267.2)  (30.7, 344.0)

Values of Predictors for New Observations

New                  Contrast
Obs  Brightness (%)       (%)
  1            50.0      25.0
```

CI: $107.4 \le \mu_{Y|x_0} \le 267.2$

PI: $30.7 \le y_0 \le 344.0$

These intervals are wider because the regressors are set at extreme values in the x space and the standard errors are greater.

12-64 Consider the stack loss data in Exercise 12-20.

(a) Calculate 95% confidence intervals on each regression coefficient.
(b) Calculate a 95% confidence interval on mean stack loss when $x_1 = 80$, $x_2 = 25$ and $x_3 = 90$.
(c) Calculate a prediction interval on stack loss for the same values of the regressors used in part (b).
(d) Calculate a 95% confidence interval and a 95% prediction interval on stack loss when $x_1 = 80$, $x_2 = 19$, and $x_3 = 93$. Compare the widths of these intervals to those calculated in parts (b) and (c). Explain any differences in widths.

(a) $t_{0.025,17} = 2.110$

$$-0.431 \le \beta_1 \le 1.00$$
$$0.519 \le \beta_2 \le 2.072$$
$$-0.482 \le \beta_3 \le 0.178$$

(b) $\hat{\mu}_{Y|x_0} = 36.023$ $\quad se(\hat{\mu}_{Y|x_0}) = 1.803$ $\quad t_{0.025,17} = 2.110$

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2,n-p} \; se(\hat{\mu}_{Y|x_0})$$
$$36.023 \pm (2.110)(1.803)$$
$$32.219 \le \mu_{Y|x_0} \le 39.827$$

(c) $\hat{y}_0 = 36.023$ $\quad se(\hat{y}_0) = 3.698$

$$36.023 \pm 2.110(3.698)$$
$$28.194 \le y_0 \le 43.852$$

(d) Prediction at $x_1 = 80$, $x_2 = 19$, $x_3 = 93$ is $\hat{\mu}_{Y|x_0} = 27.795$

CI: $21.030 \le \mu_{Y|x_0} \le 34.559$

PI: $18.173 \le y_0 \le 37.417$

12-65 Consider the NFL data in Exercise 12-21.

(a) Find 95% confidence intervals on the regression coefficients.
(b) What is the estimated standard error of $\hat{\mu}_{Y|x_0}$ when the percentage of completions is 60%, the percentage of TDs is 4%, and the percentage of interceptions is 3%.
(c) Find a 95% confidence interval on the mean rating when the percentage of completions is 60%, the percentage of TDs is 4%, and the percentage of interceptions is 3%.

(a) The computer output follows. The output is used to obtain estimates of the coefficients and standard errors. The confidence intervals for the coefficients are computed from

$\hat{\beta} - t_{0.025,28} se(\hat{\beta}) \le \beta \le \hat{\beta} + t_{0.025,28} se(\hat{\beta})$.

From the t table, $t_{0.025,\,28} = 2.048$. The confidence intervals for the β 's are

$$\begin{bmatrix} -9.052 \\ 0.999 \\ 3.807 \\ -4.808 \end{bmatrix} \le \beta \le \begin{bmatrix} 15.024 \\ 1.398 \\ 5.384 \\ -2.817 \end{bmatrix}$$

(b) From the computer output $\hat{\mu}_{Y|x_0} = 81.85, se(\hat{\mu}_{Y|x_0}) = \sqrt{V(\hat{\mu}_{Y|x_0})} = \sigma\sqrt{x_0'(X'X)^{-1}x_0} = 0.43$

(c) $\hat{\mu}_{Y|x_0} - t_{0.025,28} se(\hat{\mu}_{Y|x_0}) \le \mu_{Y|x_0} \le \hat{\mu}_{Y|x_0} + t_{0.025,28} se(\hat{\mu}_{Y|x_0})$,

$81.85 - 2.048(0.43) \le \mu_{Y|x_0} \le 81.85 + 2.048(0.43)$

This approximately equals (because of round off) the confidence interval `80.965, 82.725)` the computer output.

Regression Analysis: Rating Pts versus Pct Comp, Pct TD, Pct Int

```
The regression equation is
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int

Predictor      Coef  SE Coef      T      P
Constant      2.986    5.877   0.51  0.615
Pct Comp    1.19857  0.09743  12.30  0.000
Pct TD       4.5956   0.3848  11.94  0.000
Pct Int     -3.8125   0.4861  -7.84  0.000

S = 2.03479   R-Sq = 95.3%   R-Sq(adj) = 94.8%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  2373.59  791.20  191.09  0.000
Residual Error  28   115.93    4.14
Total           31  2489.52

Source    DF   Seq SS
Pct Comp   1  1614.43
Pct TD     1   504.49
Pct Int    1   254.67

Unusual Observations

Obs  Pct Comp  Rating Pts     Fit  SE Fit  Residual  St Resid
 11      61.1      87.700  83.691   0.371     4.009     2.00R
 18      59.4      84.700  79.668   0.430     5.032     2.53R
 21      65.7      81.000  85.020   1.028    -4.020    -2.29R
 31      59.4      70.000  75.141   0.719    -5.141    -2.70R

R denotes an observation with a large standardized residual.

Predicted Values for New Observations
```

```
New
Obs     Fit  SE Fit       95% CI            95% PI
  1  81.845   0.430  (80.965, 82.725)  (77.585, 86.104)
```

12-66 Consider the heat-treating data from Exercise 12-14.

(a) Find 95% confidence intervals on the regression coefficients.
(b) Find a 95% confidence interval on mean PITCH when TEMP = 1650, SOAKTIME = 1.00, SOAKPCT = 1.10, DIFFTIME = 1.00, and DIFFPCT = 0.80.
(c) Fit a model to PITCH using regressors x_1 = SOAKTIME × SOAKPCT and x_2 = DIFFTIME × DIFFPCT. Using the model with regressors x_1 and x_2, find a 95% confidence interval on mean PITCH when SOAKTIME = 1.00, SOAKPCT = 1.10, DIFFTIME = 1.00, and DIFFPCT = 0.80.
(d) Compare the length of this confidence interval with the length of the confidence interval on mean PITCH at the same point from part (b), which used an additive model in SOAKTIME, SOAKPCT, DIFFTIME, and DIFFPCT. Which confidence interval is shorter? Does this tell you anything about which model is preferable?

(a) $-0.000042 \le \beta_1 \le 0.000099$

$$0.00196 \le \beta_2 \le 0.00268$$
$$-0.01504 \le \beta_3 \le 0.00898$$
$$0.00597 \le \beta_4 \le 0.010979$$
$$-0.01897 \le \beta_5 \le 0.01424$$

(b) $\hat{\mu}_{Y|x_0} = 0.022466 \quad se(\hat{\mu}_{Y|x_0}) = 0.000595 \quad t_{.025,26} = 2.056$

$$0.0220086 \pm (2.056)(0.000595)$$
$$0.0212 \le \mu_{Y|x_0} \le 0.0237$$

(c) $\hat{\mu}_{Y|x_0} = 0.0171 \quad se(\hat{\mu}_{Y|x_0}) = 0.000548 \quad t_{.025,29} = 2.045$

$$0.0171 \pm (2.045)(0.000548)$$
$$0.0159 \le \mu_{Y|x_0} \le 0.0183$$

(d) Width = 2.2 E-3 and width = 2.4 E-3
The interaction model has a shorter confidence interval. Yes, this suggests the interaction model is preferable.

12-67 Consider the gasoline mileage data in Exercise 12-11.

(a) Find 99% confidence intervals on the regression coefficients.
(b) Find a 99% confidence interval on the mean of Y for the regressor values in the first row of data.
(c) Fit a new regression model to these data using *cid*, *etw*, and *axle* as the regressors. Find 99% confidence intervals on the regression coefficients in this new model.
(d) Compare the lengths of the confidence intervals from part (c) with those found in part (a). Which intervals are longer? Does this offer any insight about which model is preferable?

(a)

$t_{.005,14}$=2.977

$$-8.658 \le \beta_0 \le 108.458$$
$$-0.08 \le \beta_2 \le 0.059$$
$$-0.05 \le \beta_3 \le 0.047$$
$$-0.006 \le \beta_7 \le 0$$
$$-4.962 \le \beta_8 \le 5.546$$
$$-7.811 \le \beta_9 \le 0.101$$
$$1.102 \le \beta_{10} \le -0.523$$

(b) $\hat{\mu}_{Y|x_0} = 29.71 \qquad se(\hat{\mu}_{Y|x_0}) = 1.395$

$$\hat{\mu}_{Y|x_0} \pm t_{.005,14} \; se(\hat{\mu}_{Y|x_0})$$
$$29.71 \pm (2.977)(1.395)$$
$$25.557 \le \mu_{Y|x_0} \le 33.863$$

(c) $\hat{y} = 61.001 - 0.0208x_2 - 0.0035x_7 - 3.457x_9$

$$t_{.005,17} = 2.898$$
$$50.855 \le \beta_0 \le 71.147$$
$$-0.036 \le \beta_2 \le -0.006$$
$$-0.006 \le \beta_7 \le -0.001$$
$$-6.485 \le \beta_9 \le -0.429$$

(d) The intervals in part c) are narrower. All of the regressors used in part c) are significant, but not all of those used in part a) are significant. The model used in part c) is preferable.

12-68 Consider the NHL data in Exercise 12-22.

(a) Find a 95% confidence interval on the regression coefficient for the variable *GF*.
(b) Fit a simple linear regression model relating the response variable to the regressor *GF*.
(c) Find a 95% confidence interval on the slope for the simple linear regression model from part (b).
(d) Compare the lengths of the two confidence intervals computed in parts (a) and (c). Which interval is shorter? Does this tell you anything about which model is preferable?

(a) From the computer output, the estimate, standard error, t statistic and P-value for the coefficient of GF are:

```
Predictor       Coef  SE Coef      T      P
GF           0.16374  0.03673   4.46  0.000
```

The 95% CI on the regression coefficient β_1 of GF is

$$\hat{\beta}_1 - t_{\alpha/2,n-p} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{\alpha/2,n-p} se(\hat{\beta}_1)$$
$$\hat{\beta}_1 - t_{0.025,15} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{0.025,15} se(\hat{\beta}_1)$$
$$0.16374 - (2.131)(0.03673) \le \hat{\beta}_1 \le 0.16374 + (2.131)(0.03673)$$
$$0.085468 \le \hat{\beta}_1 \le 0.242012$$

(b) Thecomputer result is shown below.

Regression Analysis: W versus GF

```
The regression equation is
```

```
W = - 8.57 + 0.212 GF

Predictor     Coef  SE Coef      T      P
Constant    -8.574    8.910  -0.96  0.344
GF         0.21228  0.03795   5.59  0.000

S = 5.02329   R-Sq = 52.8%   R-Sq(adj) = 51.1%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   789.46  789.46  31.29  0.000
Residual Error  28   706.54   25.23
Total           29  1496.00
```

$$\hat{y} = -8.57 + 0.212 x_1$$

(c) The 95% CI on the regression coefficient β_1 of GF is

$$\hat{\beta}_1 - t_{\alpha/2,n-p} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{\alpha/2,n-p} se(\hat{\beta}_1)$$

$$\hat{\beta}_1 - t_{0.025,28} se(\hat{\beta}_1) \le \hat{\beta}_1 \le \hat{\beta}_1 + t_{0.025,28} se(\hat{\beta}_1)$$

$$0.21228 - (2.048)(0.03795) \le \hat{\beta}_1 \le 0.21228 + (2.048)(0.03795)$$

$$0.134558 \le \hat{\beta}_1 \le 0.290002$$

(d) The simple linear regression model has the narrower interval. Obviously there are extraneous variables in the model from part a). The shorter interval is an initial indicator that the original model with all variables might be improved. One might expect there are other good predictors in the model from part a), only one of which is included in the model of part b).

Section 12-5

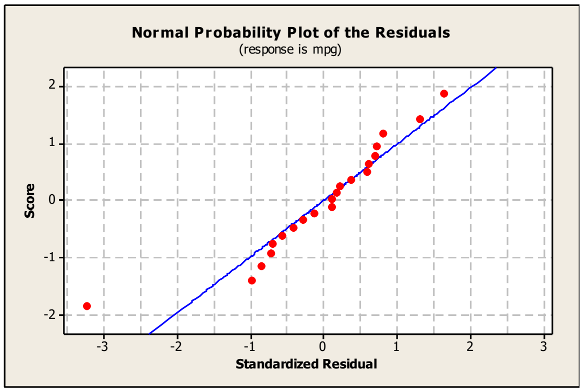
12-69 Consider the gasoline mileage data in Exercise 12-11.

(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals and comment on the normality assumption.
(c) Plot residuals versus $\hat{y}$ and versus each regressor. Discuss these residual plots.
(d) Calculate Cook's distance for the observations in this data set. Are any observations influential?

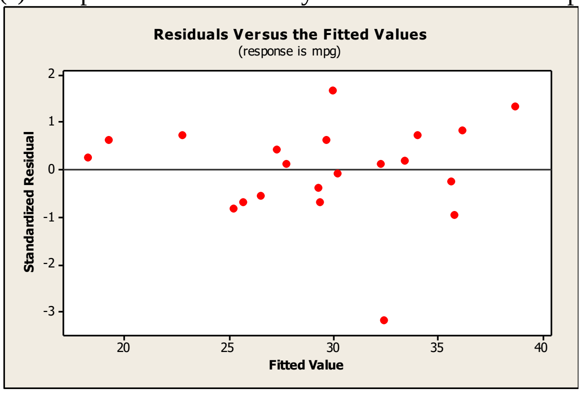
(a)

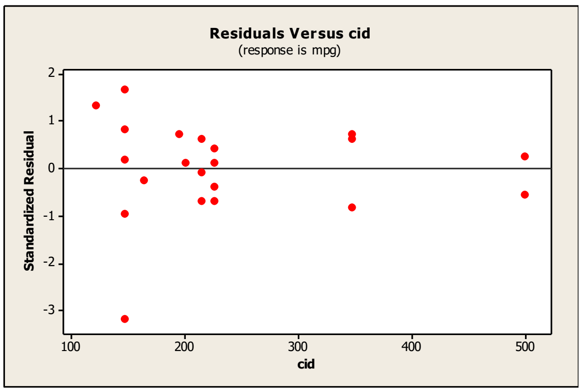
```
The regression equation is
mpg = 49.9 - 0.0104 cid - 0.0012 rhp - 0.00324 etw + 0.29 cmp - 3.86 axle
      + 0.190 n/v

Predictor        Coef    SE Coef      T      P
Constant        49.90      19.67   2.54  0.024
cid          -0.01045    0.02338  -0.45  0.662
rhp          -0.00120    0.01631  -0.07  0.942
etw        -0.0032364  0.0009459  -3.42  0.004
cmp             0.292      1.765   0.17  0.871
axle           -3.855      1.329  -2.90  0.012
n/v            0.1897     0.2730   0.69  0.498

S = 2.22830   R-Sq = 89.3%   R-Sq(adj) = 84.8%
```

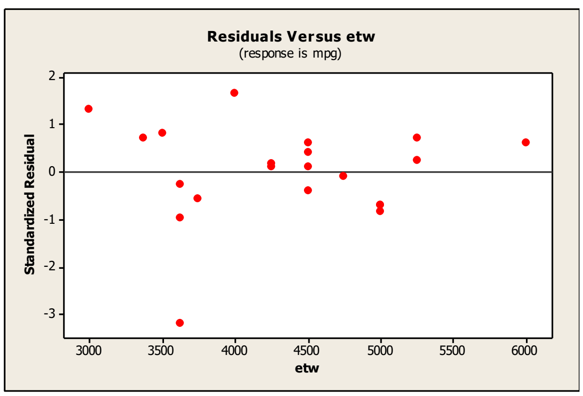
(b) There appears to be an outlier. Otherwise, the normality assumption is not violated.

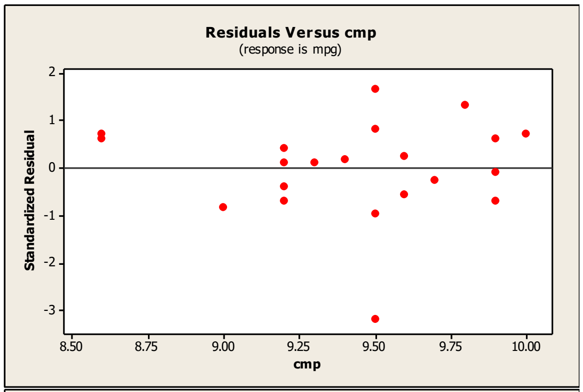


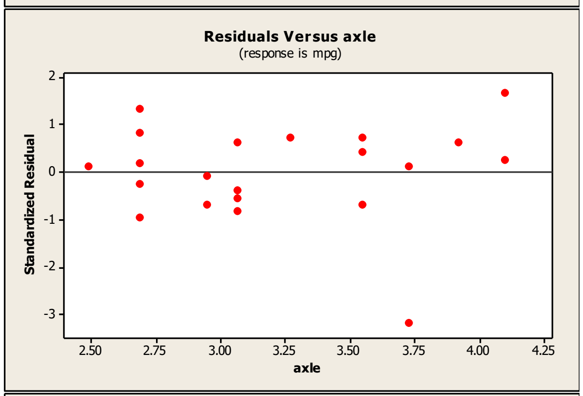
(c) The plots do not show any violations of the assumptions.

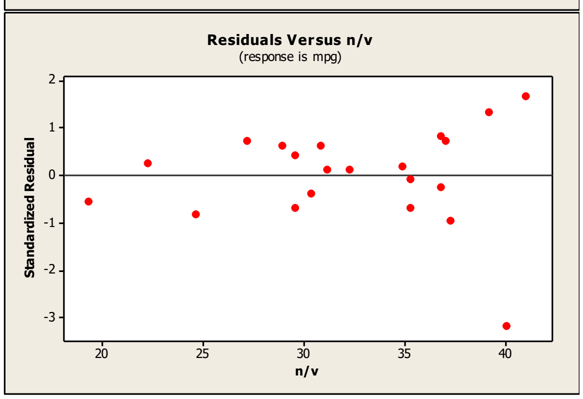












(d)
0.036216, 0.000627, 0.041684, 0.008518, 0.026788, 0.040384, 0.003136, 0.196794, 0.267746, 0.000659, 0.075126, 0.000690, 0.041624, 0.070352, 0.008565, 0.051335, 0.001813, 0.019352, 0.000812, 0.098405, 0.574353
None of the values is greater than 1 so none of the observations are influential.

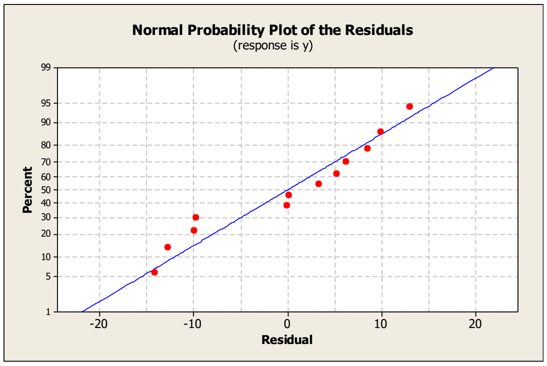
12-70 Consider the electric power consumption data in Exercise 12-10.

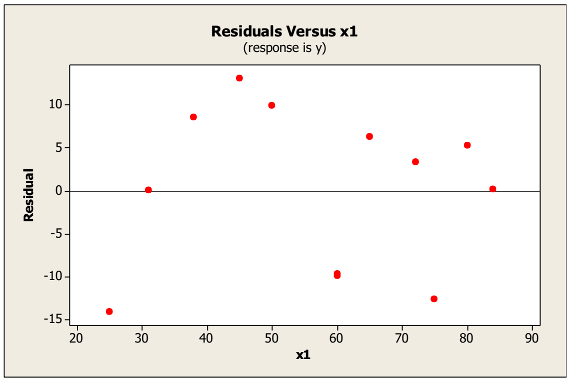
(a) Calculate R^2 for this model. Interpret this quantity.
(b) Plot the residuals versus $\hat{y}$ and versus each regressor. Interpret this plot.
(c) Construct a normal probability plot of the residuals and comment on the normality assumption.

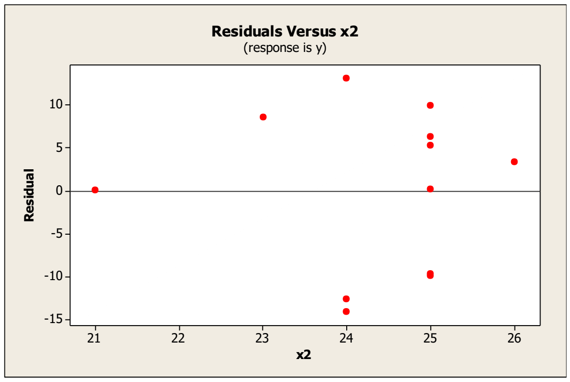
(a) $R^2 = 0.852$

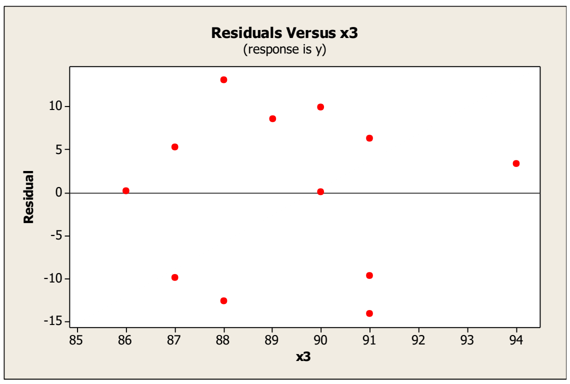
(b) The residual plots look reasonable. There is some increase in variability at the middle of the predicted values.

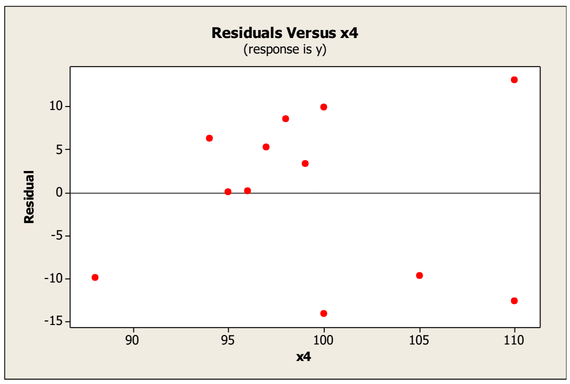
(c) Normality assumption is reasonable. The residual plots appear reasonable too.











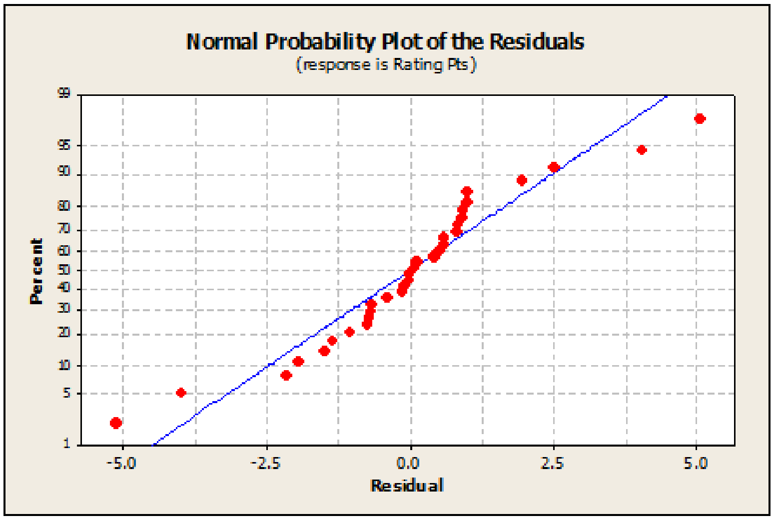
12-71 Consider the regression model for the NFL data in Exercise 12-21.

(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
(c) Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
(d) Are there any influential points in these data?

(a) The computer output follows. The proportion of total variability explained by this model is:

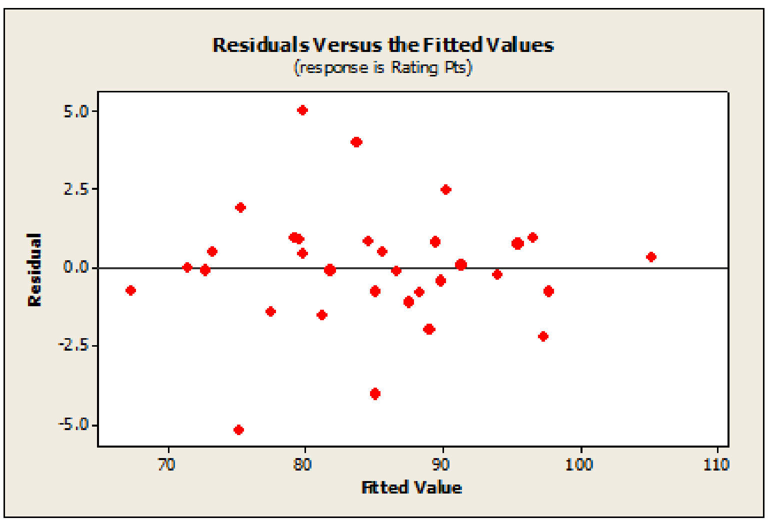
$$R^2 = \frac{SS_R}{SS_T} = \frac{2373.59}{2489.52} = 0.95$$

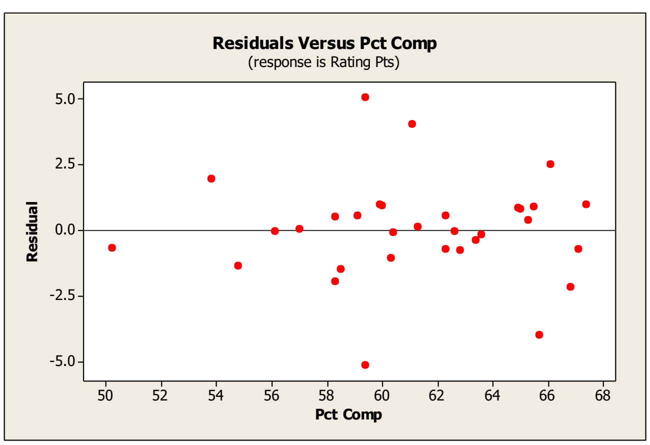
(b) Normal Probability Plot: Some moderate, but not severe, departures from normality are indicated.

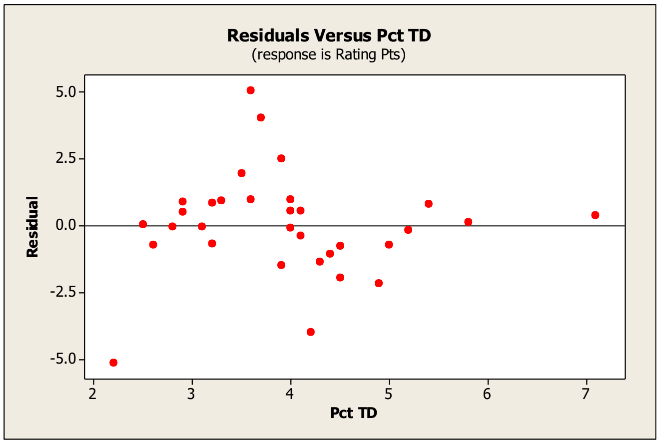


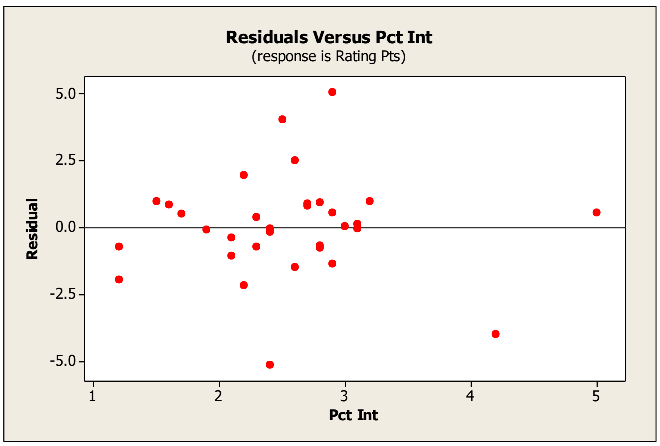
(c) Plot the residuals versus fitted value and versus each regressor.

There is no obvious model failure in the plot of fitted values versus residuals. There is a modest increase in variability in the middle range of fitted values. The residual versus PctTD shows some non-random patterns. Possibly a non-linear term would benefit the model.

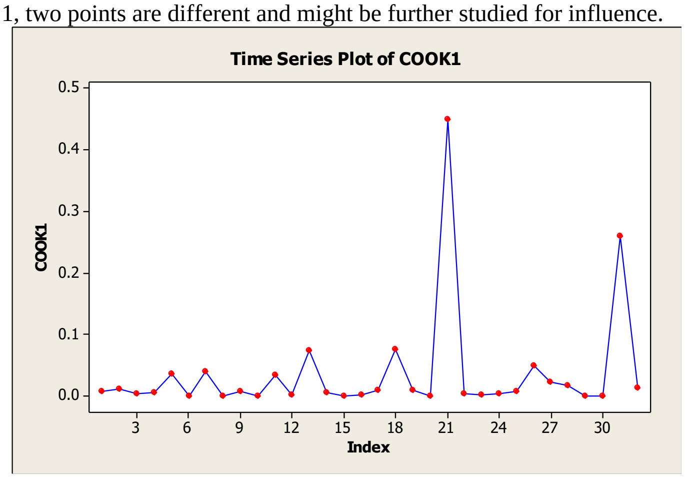








(d) A plot of Cook's distance measures follows. Although no points exceed the usual criterion of distance greater than 1, two points are different and might be further studied for influence.



Regression Analysis: Rating Pts versus Pct Comp, Pct TD, Pct Int

```
The regression equation is
Rating Pts = 2.99 + 1.20 Pct Comp + 4.60 Pct TD - 3.81 Pct Int

Predictor      Coef  SE Coef      T      P
Constant      2.986    5.877   0.51  0.615
Pct Comp    1.19857  0.09743  12.30  0.000
```

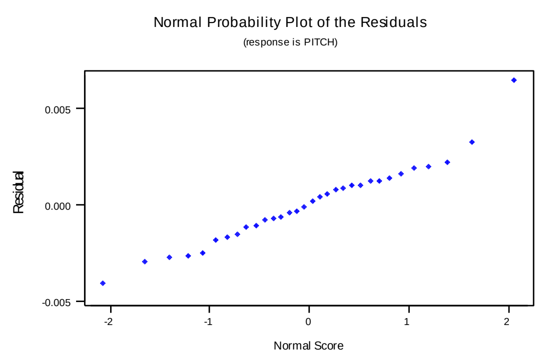
```
Pct TD       4.5956   0.3848  11.94  0.000
Pct Int     -3.8125   0.4861  -7.84  0.000

S = 2.03479   R-Sq = 95.3%   R-Sq(adj) = 94.8%

Analysis of Variance

Source           DF       SS      MS       F      P
Regression        3  2373.59  791.20  191.09  0.000
Residual Error   28   115.93    4.14
Total            31  2489.52

Source    DF   Seq SS
Pct Comp   1  1614.43
Pct TD     1   504.49
Pct Int    1   254.67

Unusual Observations

Obs  Pct Comp  Rating Pts     Fit  SE Fit  Residual  St Resid
 11      61.1      87.700  83.691   0.371     4.009     2.00R
 18      59.4      84.700  79.668   0.430     5.032     2.53R
 21      65.7      81.000  85.020   1.028    -4.020    -2.29R
 31      59.4      70.000  75.141   0.719    -5.141    -2.70R
R denotes an observation with a large standardized residual.
```

12-72 Consider the regression model for the heat-treating data in Exercise 12-14.

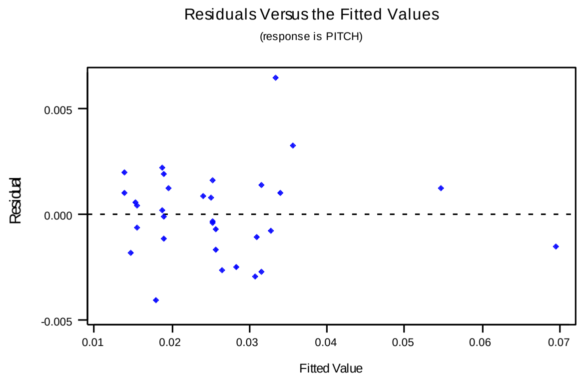
(a) Calculate the percent of variability explained by this model.
(b) Construct a normal probability plot for the residuals. Comment on the normality assumption.
(c) Plot the residuals versus $\hat{y}$ and interpret the display.
(d) Calculate Cook's distance for each observation and provide an interpretation of this statistic.

(a) R^2= 0.969

(b) Normality is acceptable



(c) Plot is acceptable.



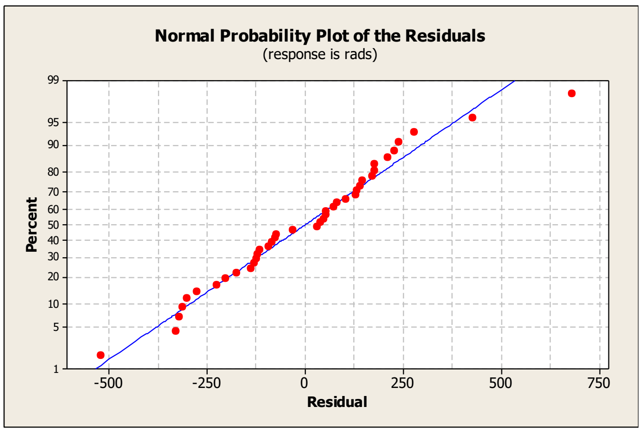
(d) Cook's distance values
0.0191 0.0003 0.0026 0.0009 0.0293 0.1112 0.1014 0.0131 0.0076 0.0004 0.0109 0.0000 0.0140 0.0039 0.0002 0.0003 0.0079 0.0022 4.5975* 0.0033 0.0058 0.14120.0161 0.0268 0.0609 0.0016 0.0029 0.3391 0.3918 0.0134 0.0088 0.5063
The 19th observation is influential

12-73 Consider the regression model fit to the X-ray inspection data in Exercise 12-15. Use rads as the response.

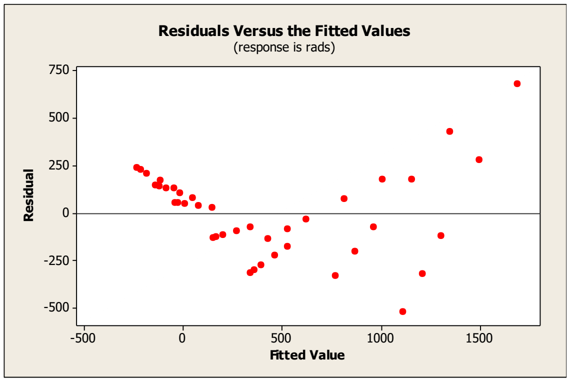
(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
(c) Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
(d) Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

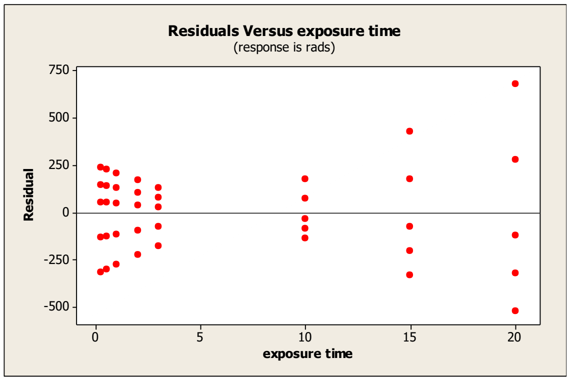
(a) $R^2 = 84.3\%$

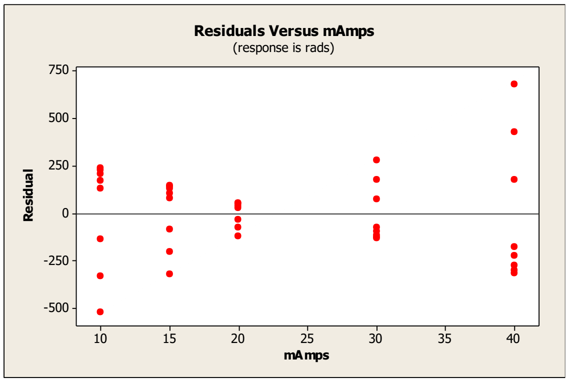
(b) Assumption of normality appears adequate.



(c) There are funnel shapes in the graphs, so the assumption of constant variance is violated. The model is inadequate.







(d) Cook's distance values

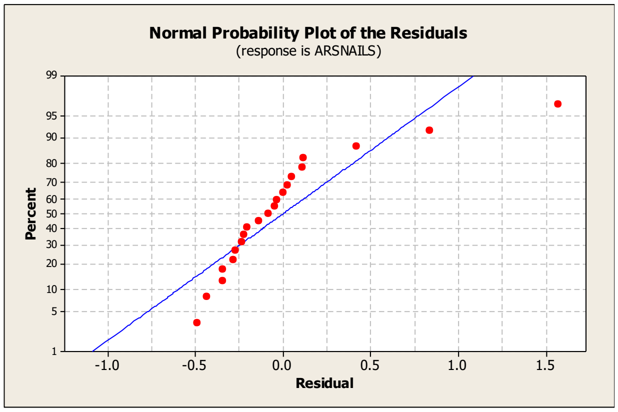
0.032728	0.029489	0.023724	0.014663	0.008279	0.008611
0.077299	0.3436	0.008489	0.007592	0.006018	0.003612
0.001985	0.002068	0.021386	0.105059	0.000926	0.000823
0.000643	0.000375	0.0002	0.000209	0.002467	0.013062
0.006095	0.005442	0.0043	0.002564	0.0014	0.001459
0.015557	0.077846	0.07828	0.070853	0.057512	0.036157
0.020725	0.021539	0.177299	0.731526		

No, none of the observations has a Cook's distance greater than 1.

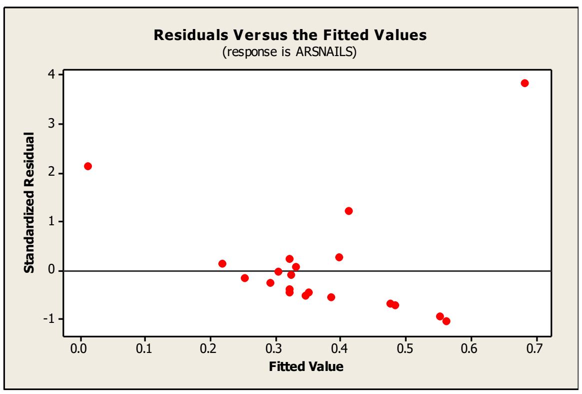
12-74 Consider the regression model fit to the arsenic data in Exercise 12-16. Use arsenic in nails as the response and age, drink use, and cook use as the regressors.
(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
(c) Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
(d) Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

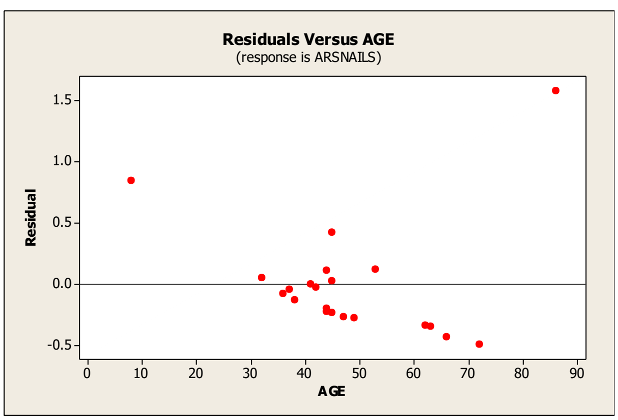
(a) $R^2 = 8.1\%$

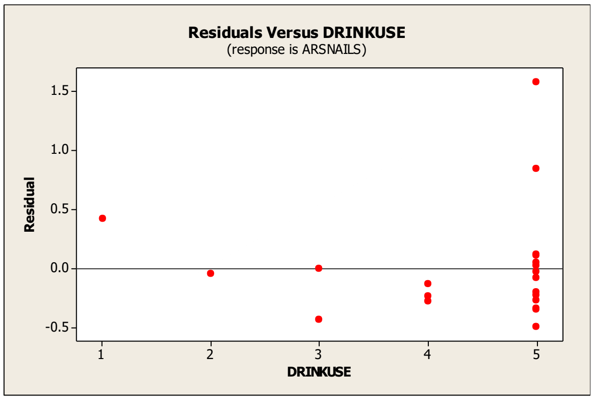
(b) Assumption of normality is not adequate.

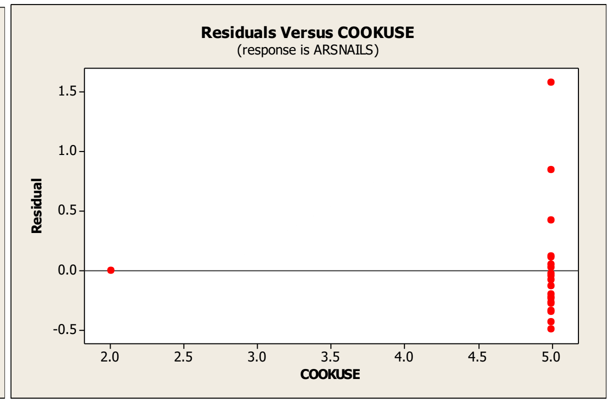


(c) The graphs indicate non-constant variance. Therefore, the model is not adequate.









(d) Cook's distance values
0.0032 0.0035 0.00386 0.05844 0.00139 0.00005 0.00524 0.00154 *infinity*
0.00496 0.05976 0.37409 0.00105 1.89094 0.68988 0.00035 0.00092 0.0155
0.00008 0.0143 0.00071

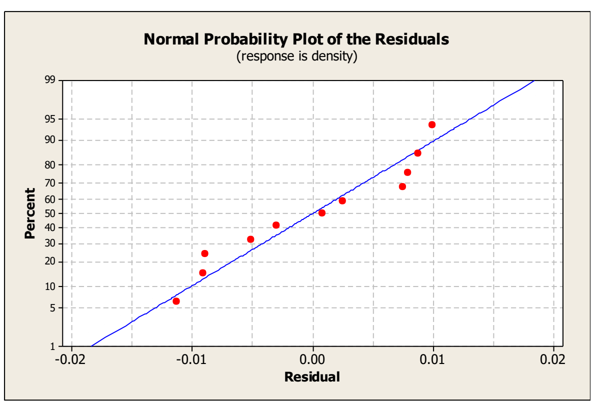
There are two influential points with Cook's distance greater than one. The entry *infinity* in the list above indicate a data point with $h_{ii} = 1$ and an undefined studentized residual.

12-75 Consider the regression model fit to the coal and limestone mixture data in Exercise 12-17. Use density as the response.

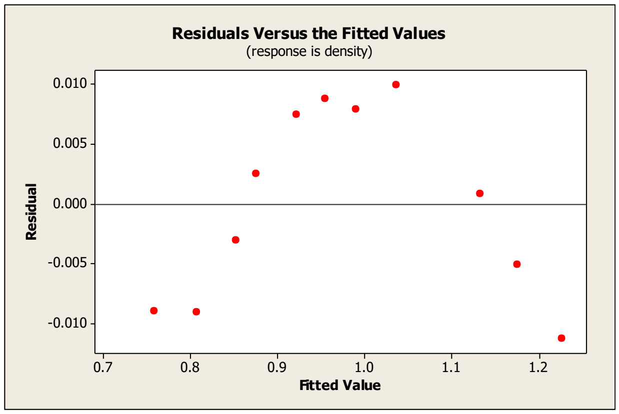
(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
(c) Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
(d) Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

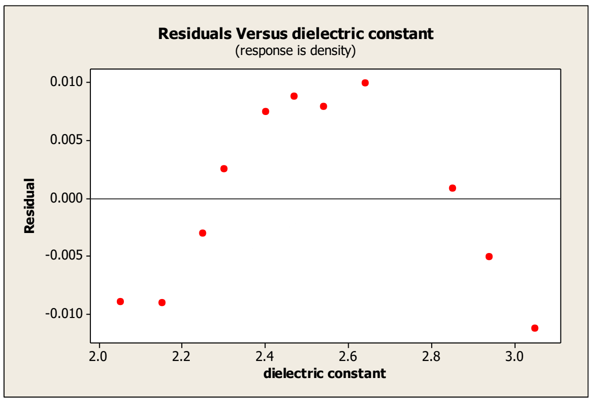
(a) $R^2 = 99.7\,\%$

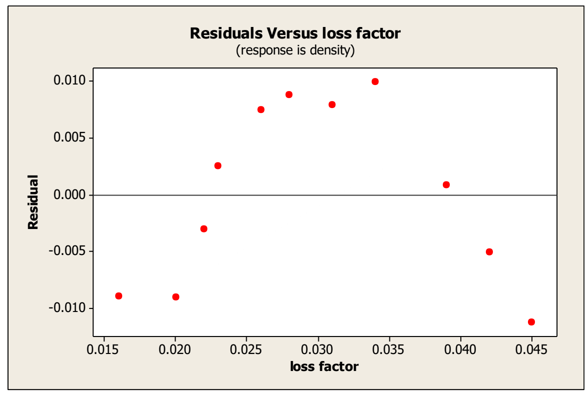
(b) Assumption of normality appears adequate.



(c) There is a non-constant variance shown in graphs. Therefore, the model is inadequate.







(d) Cook's distance values

0.255007	0.692448	0.008618	0.011784	0.058551	0.077203
0.10971	0.287682	0.001337	0.054084	0.485253	

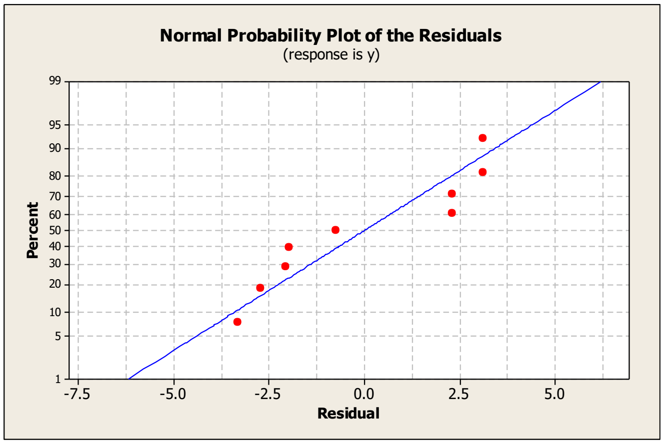
No, none of the observations has a Cook's distance greater than 1.

12-76 Consider the regression model fit to the nisin extraction data in Exercise 12-18.

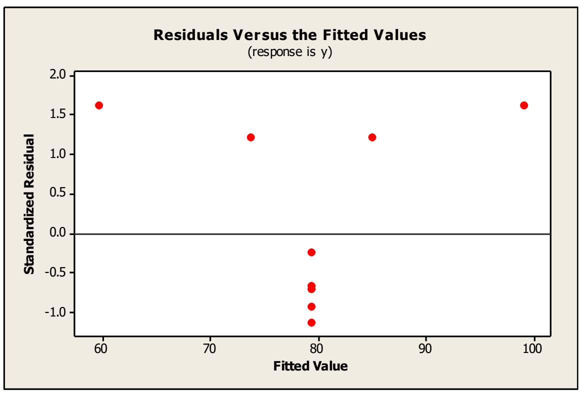
(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
(c) Plot the residuals versus ŷ and versus each regressor, and comment on model adequacy.
(d) Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

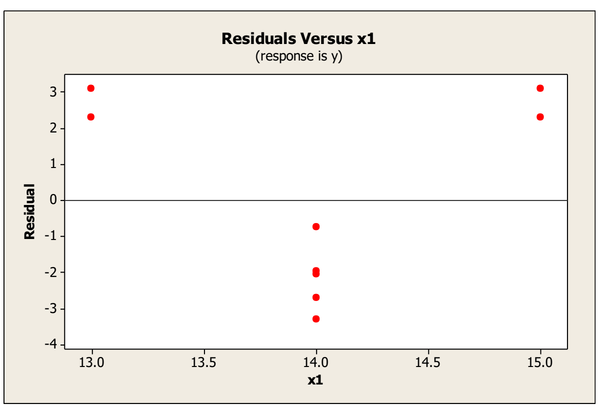
(a) $R^2 = 93.7\,\%$

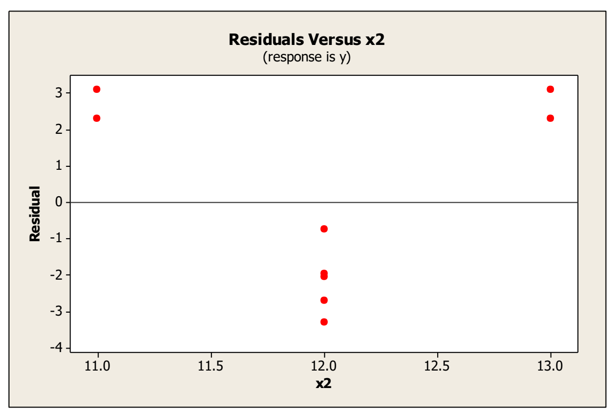
(b) The normal assumption appears inadequate



(c) The constant variance assumption is not invalid.







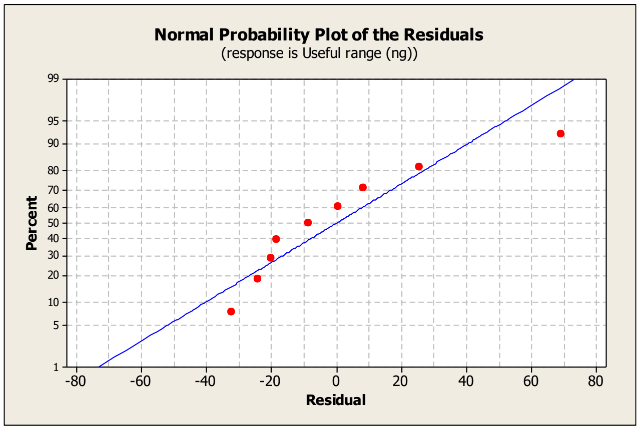
(d) Cook's distance values
1.36736 0.7536 0.7536 1.36736 0.0542 0.01917 0.03646 0.02097 0.00282
There are two influential points with Cook's distances greater than 1.

12-77 Consider the regression model fit to the gray range modulation data in Exercise 12-19. Use the useful range as the response.

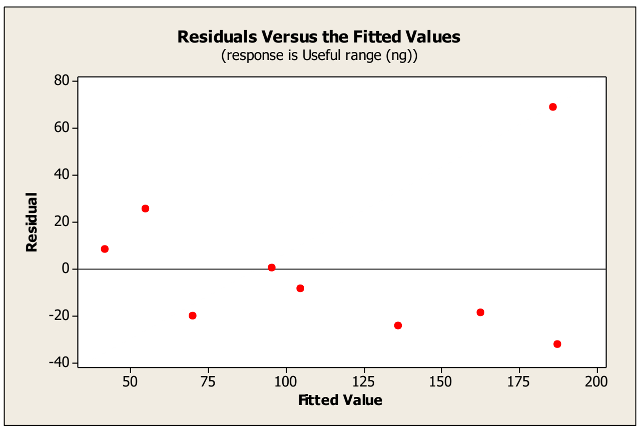
(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
(c) Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
(d) Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

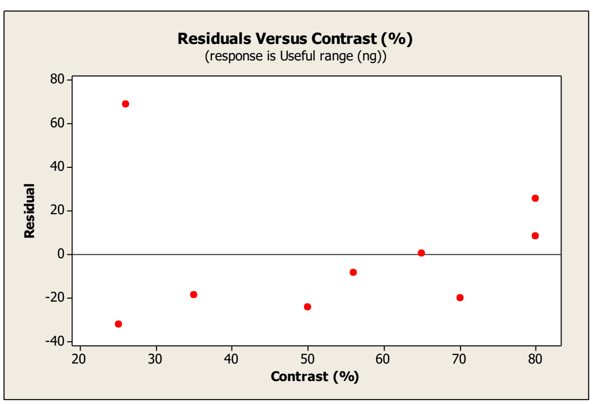
(a) $R^2 = 75.6\,\%$

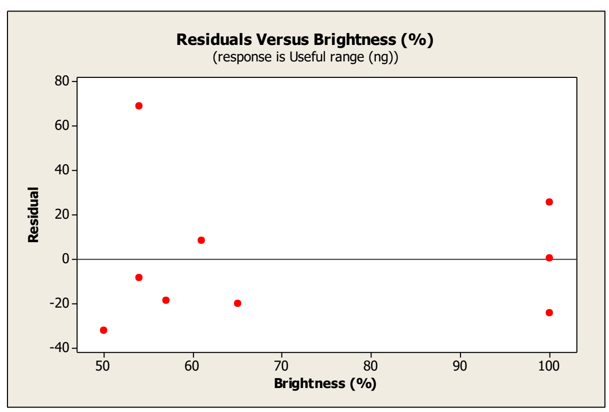
(b) Assumption of normality appears adequate.



(c) Assumption of constant variance is a possible concern. One point is a concern as a possible outlier.







(d) Cook's distance values

0.006827	0.032075	0.045342	0.213024	0.000075
0.154825	0.220637	0.030276	0.859916	

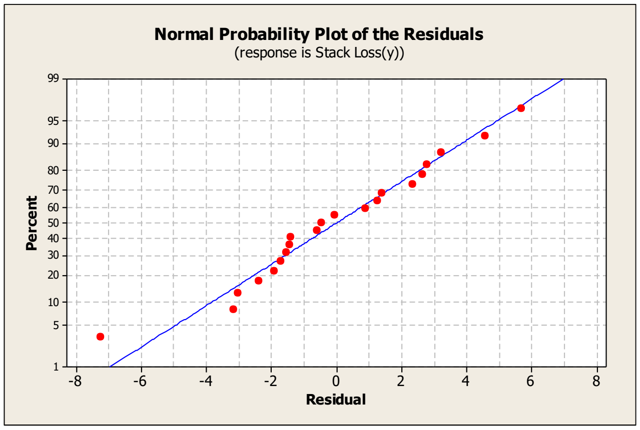
No, none of the observations has a Cook's distance greater than 1.

12-78 Consider the stack loss data in Exercise 12-20.

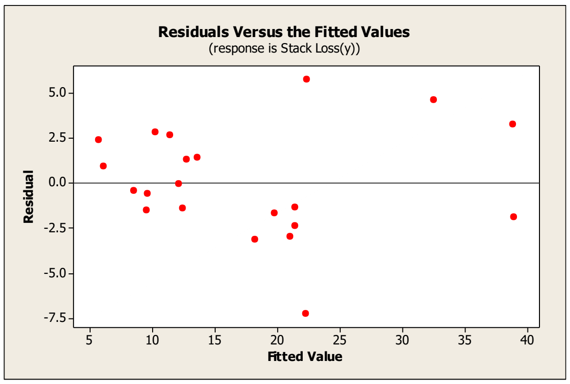
(a) What proportion of total variability is explained by this model?
(b) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?
(c) Plot the residuals versus $\hat{y}$ and versus each regressor, and comment on model adequacy.
(d) Calculate Cook's distance for the observations in this data set. Are there any influential points in these data?

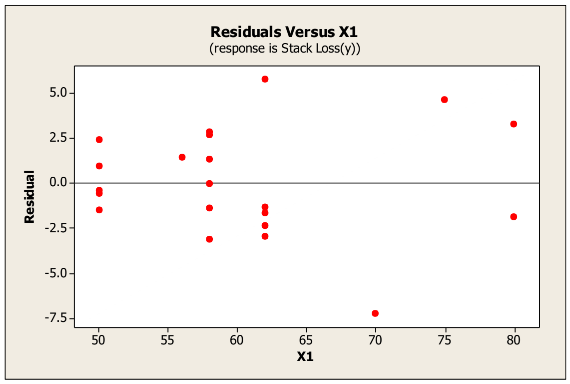
(a) $R^2 = 91.4\,\%$

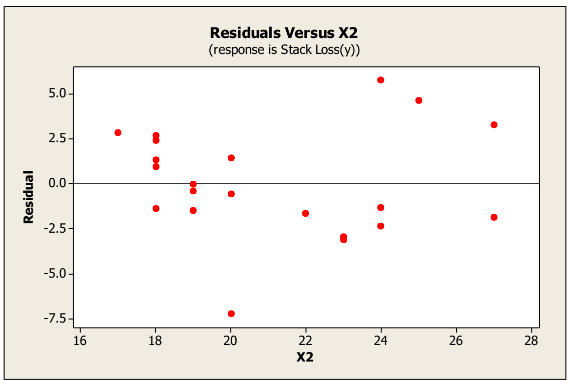
(b) Assumption of normality appears adequate.

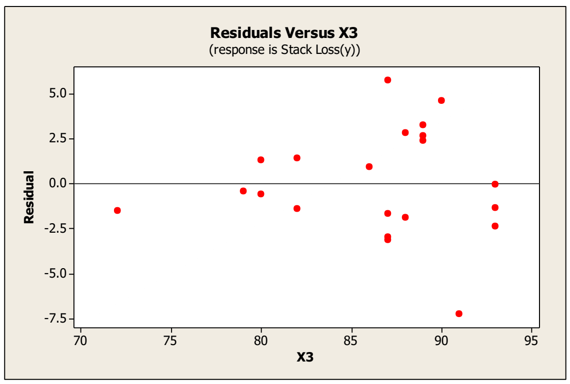


(c) Assumption of constant variance appears reasonable









(d) Cook's distance values

0.15371	0.059683	0.126414	0.130542	0.004048	0.019565
0.048802	0.016502	0.044556	0.01193	0.035866	0.065066
0.010765	0.00002	0.038516	0.003379	0.065473	0.001122
0.002179	0.004492	0.692			

No, none of the observations has a Cook's distance greater than 1.

12-79 Consider the bearing wear data in Exercise 12-23.

(a) Find the value of R^2 when the model uses the regressors x_1 and x_2.
(b) What happens to the value of R^2 when an interaction term x_1x_2 is added to the model? Does this necessarily imply that adding the interaction term is a good idea?

(a) $R^2 = 0.985$

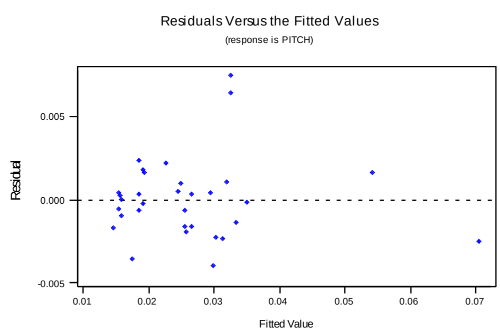
(b) $R^2 = 0.99$
R^2 increases with addition of interaction term. No, adding an additional regressor will always increase R^2.

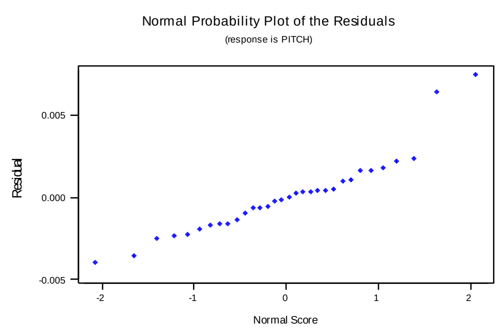
12-80 Fit a model to the response PITCH in the heat-treating data of Exercise 12-14 using new regressors x_1 = SOAKTIME × SOAKPCT and x_2 = DIFFTIME × DIFFPCT.

(a) Calculate the R^2 for this model and compare it to the value of R^2 from the original model in Exercise 12-14. Does this provide some information about which model is preferable?
(b)Plot the residuals from this model versus $\hat{y}$ and on a normal probability scale. Comment on model adequacy.
(c) Find the values of Cook's distance measure. Are any observations unusually influential?

(a) $R^2 = 0.955$. Yes, the R^2 using these two regressors is nearly as large as the R^2 from the model with five regressors.

(b) Normality is acceptable, but there is some indication of outliers.



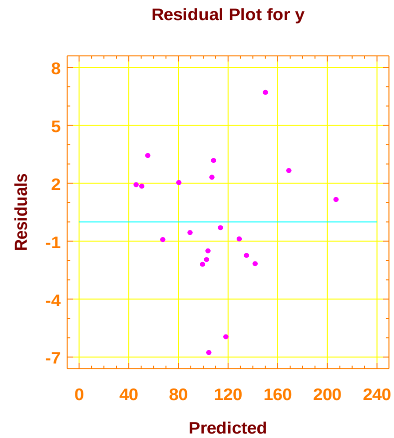


(c) Cook's distance values
0.0202 0.0008 0.0021 0.0003 0.0050 0.0000 0.0506 0.0175 0.0015 0.0003 0.0087
0.0001 0.0072 0.0126 0.0004 0.0021 0.0051 0.0007 0.0282 0.0072 0.0004 0.1566
0.0267 0.0006 0.0189 0.0179 0.0055 0.1141 0.1520 0.0001 0.0759 2.3550
The last observation is very influential

12-81 Consider the semiconductor HFE data in Exercise 12-13.

(a) Plot the residuals from this model versus $\hat{y}$. Comment on the information in this plot.
(b) What is the value of R^2 for this model?
(c) Refit the model using log HFE as the response variable.
(d) Plot the residuals versus predicted log HFE for the model in part (c). Does this give any information about which model is preferable?
(e) Plot the residuals from the model in part (d) versus the regressor x_3. Comment on this plot.
(f) Refit the model to log HFE using x_1, x_2, and $1 / x_3$, as the regressors. Comment on the effect of this change in the model.

(a) There is some indication of nonconstant variance since the residuals appear to "fan out" with increasing values of y.



(b)

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	30531.5	3	10177.2	840.546	.0000
Error	193.725	16	12.1078		
Total (Corr.)	30725.2	19			

```
R-squared = 0.993695                               Stnd. error of est. = 3.47963
R-squared (Adj. for d.f.) = 0.992513            Durbin-Watson statistic = 1.77758
```

$R^2 = 0.9937$ or 99.37 %;

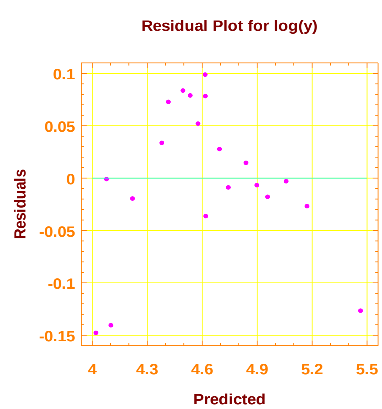
$R^2_{Adj} = 0.9925$ or 99.25%;

(c)

```
                      Model fitting results for: log(y)
--------------------------------------------------------------------------------
Independent variable            coefficient  std. error     t-value   sig.level
CONSTANT                            6.22489    1.124522      5.5356      0.0000
x1                                 -0.16647    0.083727     -1.9882      0.0642
x2                                -0.000228    0.005079     -0.0448      0.9648
x3                                 0.157312    0.029752      5.2875      0.0001
--------------------------------------------------------------------------------
R-SQ. (ADJ.) = 0.9574  SE=       0.078919  MAE=       0.053775  DurbWat=  2.031
Previously:    0.0000            0.000000             0.000000            0.000
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

$$\hat{y}^* = 6.22489 - 0.16647x_1 - 0.000228x_2 + 0.157312x_3$$

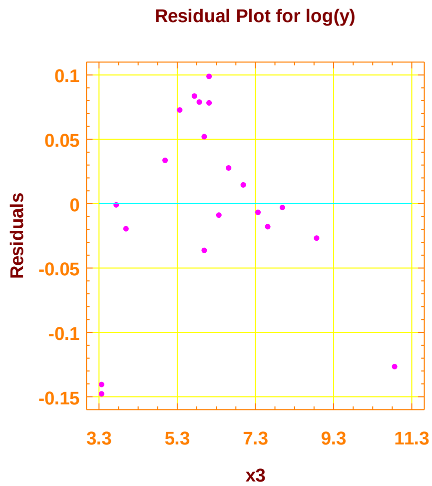
(d)



Plot exhibits curvature

There is curvature in the plot. The plot does not give much more information as to which model is preferable.

(e)

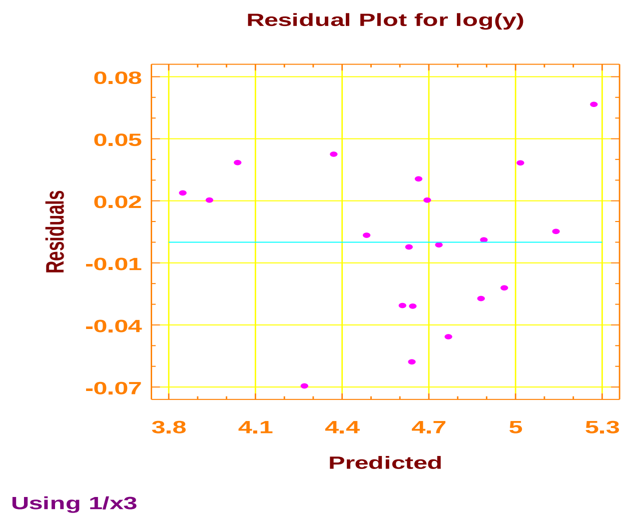


Plot exhibits curvature

Variance does not appear constant. Curvature is evident.

(f)

```
                     Model fitting results for: log(y)
Independent variable             coefficient  std. error     t-value   sig.level
CONSTANT                            6.222045    0.547157     11.3716      0.0000
x1                                 -0.198597    0.034022     -5.8374      0.0000
x2                                  0.009724    0.001864      5.2180      0.0001
1/x3                               -4.436229    0.351293    -12.6283      0.0000
--------------------------------------------------------------------------------
R-SQ. (ADJ.) = 0.9893  SE=       0.039499  MAE=       0.028896  DurbWat=  1.869


                  Analysis of Variance for the Full Regression
Source                 Sum of Squares     DF    Mean Square      F-Ratio   P-value
Model                         2.75054      3      0.916847       587.649     .0000
Error                       0.0249631     16    0.00156020
Total (Corr.)                 2.77550     19
R-squared = 0.991006                               Stnd. error of est. = 0.0394993
R-squared (Adj. for d.f.) = 0.98932              Durbin-Watson statistic = 1.86891
```



The residual plot indicates better conformance to assumptions.

Curvature is removed when using $1/x_3$ as the regressor instead of x_3, and the log of the response data.

12-82 Consider the regression model for the NHL data from Exercise 12-22.

(a) Fit a model using as the only regressor.
(b) How much variability is explained by this model?
(c) Plot the residuals versus $\hat{y}$ and comment on model adequacy.
(d) Plot the residuals from part (a) versus *PPGF*, the points scored while in power play. Does this indicate that the model would be better if this variable were included?

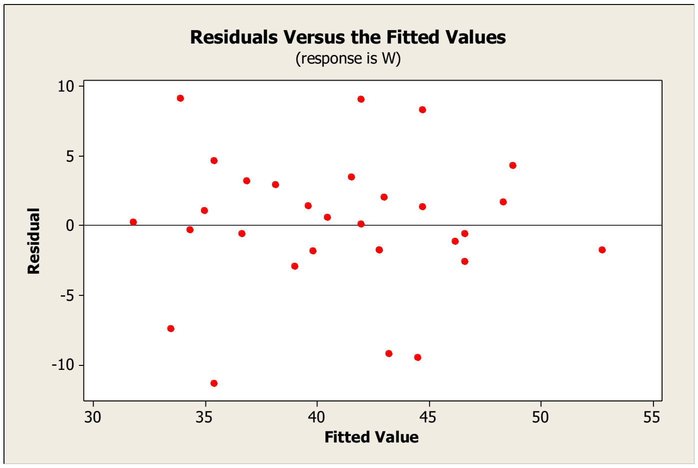
(a) The computer output is shown below.

Regression Analysis: W versus GF

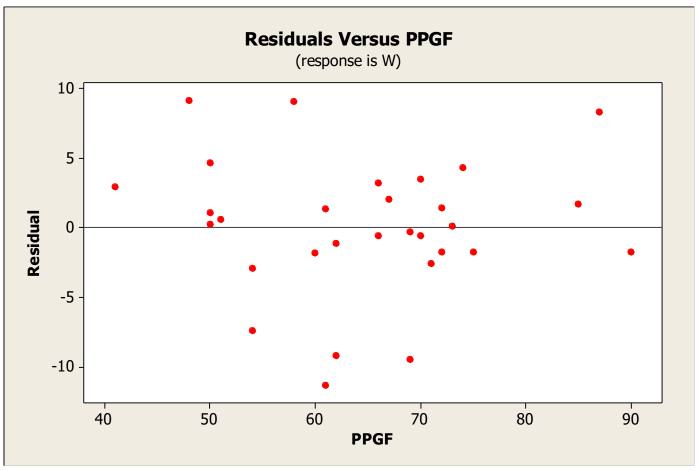
```
The regression equation is
W = - 8.57 + 0.212 GF

Predictor      Coef  SE Coef      T      P
Constant     -8.574    8.910  -0.96  0.344
GF          0.21228  0.03795   5.59  0.000

S = 5.02329   R-Sq = 52.8%   R-Sq(adj) = 51.1%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   789.46  789.46  31.29  0.000
Residual Error  28   706.54   25.23
Total           29  1496.00
```

(b) `R-Sq = 52.8%`

(c) Model appears adequate.



(d) No, the residuals do not seem to be related to PPGF. Because there is no pattern evident in the plot, it does not seem that this variable would contribute significantly to the model.



12-83 The diagonal elements of the hat matrix are often used to denote **leverage**—that is, a point that is unusual in its location in the *x*-space and that may be influential. Generally, the *i*th point is called a **leverage point** if its hat diagonal h_{ii} exceeds $2p / n$, which is twice the average size of all the hat diagonals. Recall that $p = k + 1$.

(a) Table 12-9 contains the hat diagonal for the wire bond pull strength data used in Example 12-5. Find the average size of these elements.
(b) Based on the preceding criterion, are there any observations that are leverage points in the data set?

(a) $p = k + 1 = 2 + 1 = 3$
Average size = p/n = 3/25 = 0.12

(b) Leverage point criteria:

$$h_{ii} > 2(p/n)$$

$$h_{ii} > 2(0.12)$$

$$h_{ii} > 0.24$$

$$h_{17,17} = 0.2593$$

$$h_{18,18} = 0.2929$$

Points 17 and 18 are leverage points.

Sections 12-6

12-84 An article entitled "A Method for Improving the Accuracy of Polynomial Regression Analysis'' in the *Journal of Quality Technology* (1971, pp. 149–155) reported the following data on y = ultimate shear strength of a rubber compound(psi) and x = cure temperature (°F).

y	770	800	840	810
x	280	284	292	295
y	735	640	590	560
x	298	305	308	315

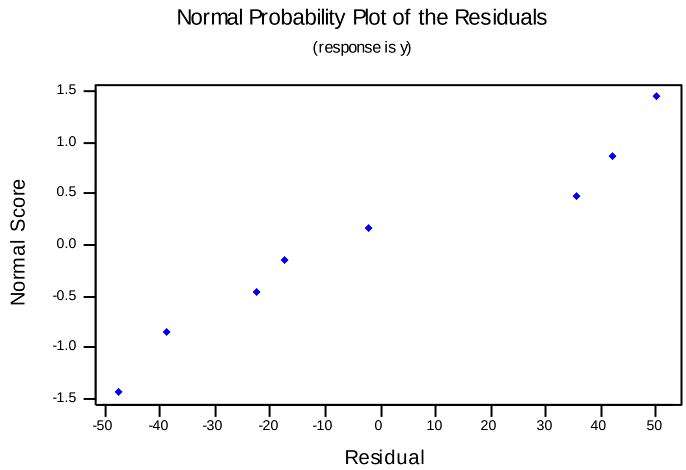
(a) Fit a second-order polynomial to these data.
(b) Test for significance of regression using α = 0.05.
(c) Test the hypothesis that $\beta_{11} = 0$ using α = 0.05.
(d) Compute the residuals from part (a) and use them to evaluate model adequacy.

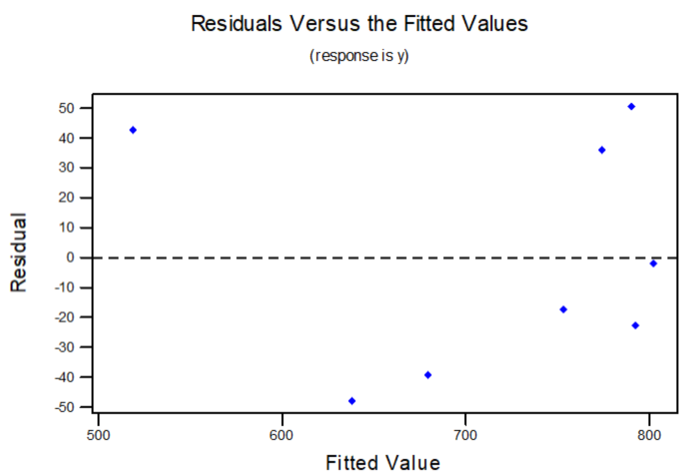
(a) $\hat{y} = -26219.15 + 189.205x - 0.33x^2$

(b) $H_0: \beta_j = 0$ for all j
$H_1: \beta_j \neq 0$ for at least one j
α = 0.05
$f_0 = 17.2045$
$f_{.05,2,5} = 5.79$
$f_0 > f_{.05,2,5}$
Reject H_0 and conclude that model is significant at α = 0.05

(c) $H_0: \beta_{11} = 0$
$H_1: \beta_{11} \neq 0$
α = 0.05
$t_0 = -2.45$
$t_{\alpha,n-p} = t_{.025,8-3} = t_{.025,5} = 2.571$
$|t_0| \not> 2.571$
Fail to reject H_0 and conclude insufficient evidence to support value of quadratic term in model at α = 0.05

(d) One residual is an outlier.
Normality assumption appears acceptable.
Residuals against fitted values is somewhat curved, but the impact of the outlier should be considered.





12-85 Consider the following data, which result from an experiment to determine the effect of x = test time in hours at a particular temperature on y = change in oil viscosity:

(a) Fit a second-order polynomial to the data.

y	–1.42	–1.39	–1.55	–1.89	–2.43
x	.25	.50	.75	1.00	1.25
y	–3.15	–4.05	–5.15	–6.43	–7.89
x	1.50	1.75	2.00	2.25	2.50

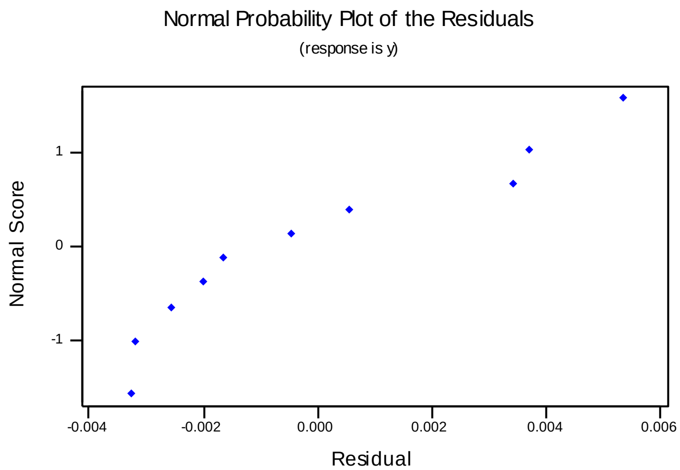
(b) Test for significance of regression using $\alpha = 0.05$.
(c) Test the hypothesis that $\beta_{11} = 0$ using $\alpha = 0.05$.
(d) Compute the residuals from part (a) and use them to evaluate model adequacy.

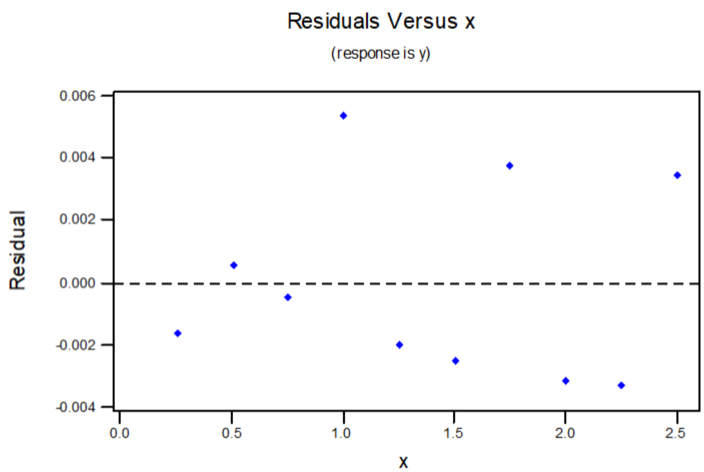
(a) $\hat{y} = -1.633 + 1.232x - 1.495x^2$

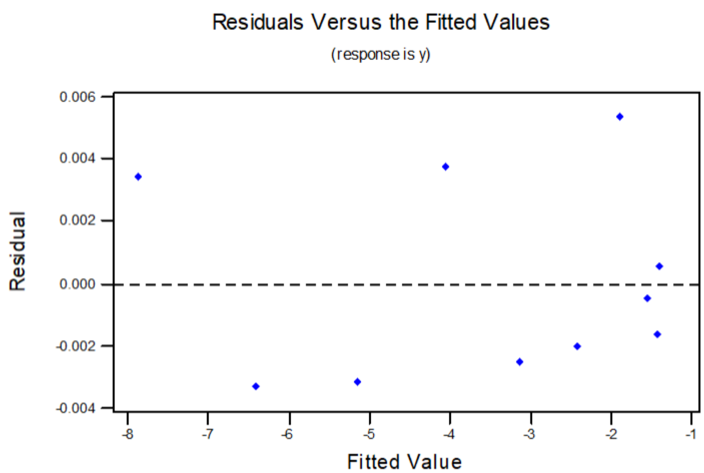
(b) $f_0 = 1858613$, reject H_0

(c) $t_0 = -601.64$, reject H_0

(d) Model is acceptable. Observation number 10 has large leverage.







12-86 The following data were collected during an experiment to determine the change in thrust efficiency (*y*, in percent) as the divergence angle of a rocket nozzle (*x*) changes:

y	24.60	24.71	23.90	39.50	39.60	57.12
x	4.0	4.0	4.0	5.0	5.0	6.0
y	67.11	67.24	67.15	77.87	80.11	84.67
x	6.5	6.5	6.75	7.0	7.1	7.3

(a) Fit a second-order model to the data.
(b) Test for significance of regression and lack of fit using $\alpha = 0.05$.
(c) Test the hypothesis that $\beta_{11} = 0$, using $\alpha = 0.05$.
(d) Plot the residuals and comment on model adequacy.
(e) Fit a cubic model, and test for the significance of the cubic term using $\alpha = 0.05$.

(a) $\hat{y} = -4.46 + 1.38x + 1.47x^2$

(b) $H_0 : \beta_j = 0$ for all j

$H_1 : \beta_j \neq 0$ for at least one j

$\alpha = 0.05$

$f_0 = 1044.99$

$f_{.05,2,9} = 4.26$

$f_0 > f_{0.05,2,9}$

Reject H_0 and conclude that the regression model is significant at $\alpha = 0.05$

(c) $H_0 : \beta_{11} = 0$

$H_1 : \beta_{11} \neq 0$ $\alpha = 0.05$

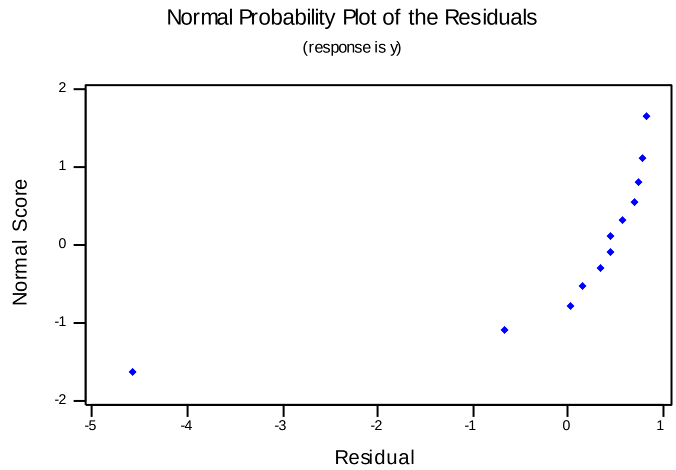
$t_0 = 2.97$

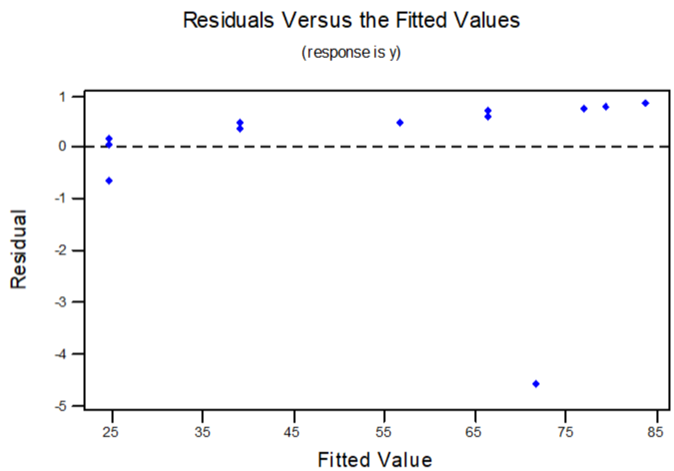
$t_{.025,9} = 2.262$

$| t_0 | > t_{0.025,9}$

Reject H_0 and conclude that β_{11} is significant at $\alpha = 0.05$

(d) Observation 9 is an extreme outlier.





(e) $\hat{y} = -87.36 + 48.01x - 7.04x^2 + 0.51x^3$

$H_0 : \beta_{33} = 0$

$H_1 : \beta_{33} \neq 0 \quad \alpha = 0.05$

$t_0 = 0.91$

$t_{.025,8} = 2.306$

$|t_0| \not> t_{0.025,8}$

Fail to reject H_0 and conclude that cubic term is not significant at $\alpha = 0.05$

12-87 An article in the *Journal of Pharmaceuticals Sciences* (1991, Vol. 80, pp. 971–977) presents data on the observed mole fraction solubility of a solute at a constant temperature and the dispersion, dipolar, and hydrogen-bonding Hansen partial solubility parameters. The data are as shown in the Table E12-13, where y is the negative logarithm of the mole fraction solubility, x_1 is the dispersion partial solubility, x_2 is the dipolar partial solubility, and x_3 is the hydrogen-bonding partial solubility.

(a) Fit the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \in .$$

(b) Test for significance of regression using $\alpha = 0.05$.

(c) Plot the residuals and comment on model adequacy.

(d) Use the extra sum of squares method to test the contribution of the second-order terms using α = 0.05.

TABLE • E12-13 Solubility Data

Observation Number	y	x_1	x_2	x_3
1	0.22200	7.3	0.0	0.0
2	0.39500	8.7	0.0	0.3
3	0.42200	8.8	0.7	1.0
4	0.43700	8.1	4.0	0.2
5	0.42800	9.0	0.5	1.0
6	0.46700	8.7	1.5	2.8
7	0.44400	9.3	2.1	1.0
8	0.37800	7.6	5.1	3.4
9	0.49400	10.0	0.0	0.3
10	0.45600	8.4	3.7	4.1
11	0.45200	9.3	3.6	2.0
12	0.11200	7.7	2.8	7.1
13	0.43200	9.8	4.2	2.0
14	0.10100	7.3	2.5	6.8
15	0.23200	8.5	2.0	6.6
16	0.30600	9.5	2.5	5.0
17	0.09230	7.4	2.8	7.8
18	0.11600	7.8	2.8	7.7
19	0.07640	7.7	3.0	8.0
20	0.43900	10.3	1.7	4.2
21	0.09440	7.8	3.3	8.5
22	0.11700	7.1	3.9	6.6
23	0.07260	7.7	4.3	9.5
24	0.04120	7.4	6.0	10.9
25	0.25100	7.3	2.0	5.2
26	0.00002	7.6	7.8	20.7

(a)
```
Predictor        Coef      SE Coef        T       P
Constant       -1.769        1.287    -1.37   0.188
x1             0.4208       0.2942     1.43   0.172
x2             0.2225       0.1307     1.70   0.108
x3           -0.12800      0.07025    -1.82   0.087
x1x2         -0.01988      0.01204    -1.65   0.118
x1x3         0.009151     0.007621     1.20   0.247
x2x3         0.002576     0.007039     0.37   0.719
x1^2         -0.01932      0.01680    -1.15   0.267
x2^2         -0.00745      0.01205    -0.62   0.545
x3^3         0.000824     0.001441     0.57   0.575

S = 0.06092     R-Sq = 91.7%     R-Sq(adj) = 87.0%

Analysis of Variance

Source            DF         SS         MS       F       P
Regression         9   0.655671   0.072852   19.63   0.000
Residual Error    16   0.059386   0.003712
Total             25   0.715057
```

$$\hat{y} = -1.769 + 0.421x_1 + 0.222x_2 - 0.128x_3 - 0.02x_{12} + 0.009x_{13}$$
$$+ 0.003x_{23} - 0.019x_1^{\,2} - 0.007x_2^{\,2} + 0.001x_3^{\,2}$$

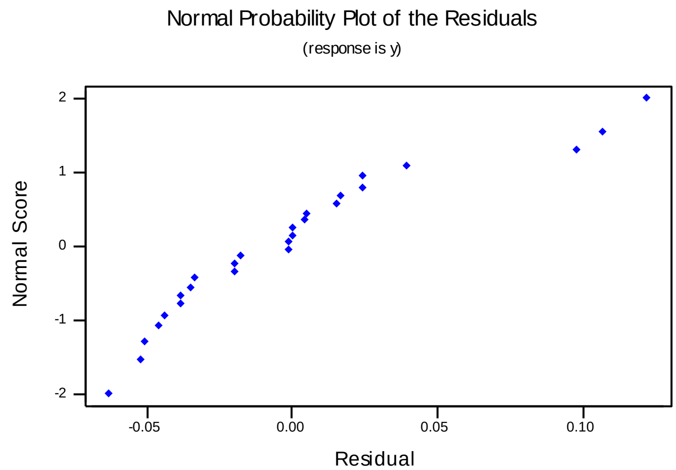
(b) H_0 all $\beta_1 = \beta_2 = \beta_3 = \cdots \beta_{23} = 0$

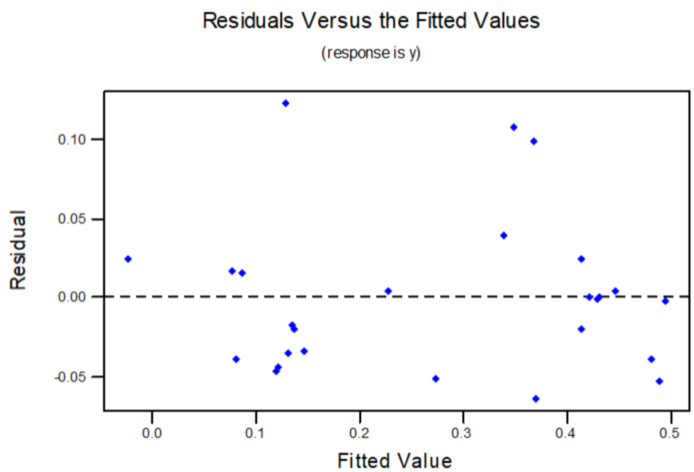
H_1 at least one $\beta_j \neq 0$

$$f_0 = 19.628$$
$$f_{.05,9,16} = 2.54$$
$$f_0 > f_{0.05,9,16}$$

Reject H_0 and conclude that the model is significant at $\alpha = 0.05$

(c) Assumptions appear to be reasonable.





(d) $H_0 : \beta_{11} = \beta_{22} = \beta_{33} = \beta_{12} = \beta_{13} = \beta_{23} = 0$

H_1 at least one $\beta \neq 0$

$$f_0 = \frac{SS_R(\beta_{11}, \beta_{22}, \beta_{33}, \beta_{12}, \beta_{13}, \beta_{23} \mid \beta_1, \beta_2, \beta_3, \beta_0)/r}{MS_E} = \frac{\frac{0.0359}{6}}{0.003712} = 1.612$$

$$f_{.05,6,16} = 2.74$$
$$f_0 \not> f_{.05,6,16}$$

Fail to reject H_0

$$SS_R(\beta_{11}\beta_{22}\beta_{33}\beta_{12}\beta_{13}\beta_{23} \mid \beta_1\beta_2\beta_3\beta_0)$$
$$= SS_R(\beta_{11}\beta_{22}\beta_{33}\beta_{12}\beta_{13}\beta_{23}\beta_1\beta_2\beta_3 \mid \beta_0) - SS_R(\beta_1\beta_2\beta_3 \mid \beta_0)$$
$$= 0.65567068 - 0.619763$$
$$= 0.0359$$

Reduced Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

12-88 Consider the arsenic concentration data in Exercise 12-16.

(a) Discuss how you would model the information about the person's sex.
(b) Fit a regression model to the arsenic in nails using age, drink use, cook use, and the person's sex as the regressors.
(c) Is there evidence that the person's sex affects arsenic in the nails? Why?

(a) Create an indicator variable for sex (e.g. 0 for male, 1 for female) and include this variable in the model.

(b)
```
The regression equation is
ARSNAILS = - 0.214 - 0.008 DRINKUSE + 0.028 COOKUSE + 0.00794 AGE + 0.167 SEXID

Predictor       Coef   SE Coef       T      P
Constant     -0.2139    0.9708   -0.22  0.828
DRINKUSE     -0.0081    0.1050   -0.08  0.940
COOKUSE       0.0276    0.1844    0.15  0.883
AGE         0.007937  0.007251    1.09  0.290
SEXID         0.1675    0.2398    0.70  0.495

S = 0.514000   R-Sq = 10.8%   R-Sq(adj) = 0.0%

where SEXID = 0 for male and 1 for female
```

(c) Because the P-value for testing $H_0: \beta_{sex} = 0$ against $H_1: \beta_{sex} \neq 0$ is 0.495, there is no evidence that the person's sex affects arsenic in the nails.

12-89 Consider the gasoline mileage data in Exercise 12-11.

(a) Discuss how you would model the information about the type of transmission in the car.
(b) Fit a regression model to the gasoline mileage using *cid*, *etw* and the type of transmission in the car as the regressors.
(c) Is there evidence that the type of transmission (L4, L5, or M6) affects gasoline mileage performance?

(a) Use indicator variable for transmission type.
There are three possible transmission types: L4, L5 and M6. So, two indicator variables could be used where x_3=1 if trns=L5, 0 otherwise and x_4=1 if trns=M6, 0 otherwise.

(b) $\hat{y} = 56.677 - 0.1457x_1 - 0.00525x_2 - 0.138x_3 - 4.179x_4$

(c) The P-value for testing $H_0: \beta_3 = 0$ is 0.919, which is not significant. However, the P-value for testing $H_0: \beta_4 = 0$ is 0.02, which is significant for values of α > 0.02. Thus, it appears that whether or not the transmission is manual affects mpg, but there is not a significant difference between the types of automatic transmission.

12-90 Consider the surface finish data in Example 12-13. Test the hypothesis that two different regression models (with different slopes and intercepts) are required to adequately model the data. Use indicator variables in answering this question.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_{12}$$
$$\hat{y} = 11.503 + 0.153x_1 - 6.094x_2 - 0.031x_{12}$$

where $x_2 = \begin{cases} 0 & \text{for tool type 302} \\ 1 & \text{for tool type 416} \end{cases}$

Test of different slopes:

$$H_0 : \beta_{12} = 0$$
$$H_1 : \beta_{12} \neq 0 \quad \alpha = 0.05$$
$$t_0 = -1.79$$
$$t_{.025,16} = 2.12$$
$$|t_0| \not> t_{0.025,16}$$

Fail to reject H_0. There is not sufficient evidence to conclude that two regression models are needed.

Test of different intercepts and slopes using extra sums of squares:

$$H_0 : \beta_2 = \beta_{12} = 0$$

H_1 at least one is not zero

$$\begin{aligned} SS(\beta_2, \beta_{12} \mid \beta_0) &= SS(\beta_1, \beta_2, \beta_{12} \mid \beta_0) - SS(\beta_1 \mid \beta_0) \\ &= 1013.35995 - 130.60910 \\ &= 882.7508 \end{aligned}$$

$$f_0 = \frac{SS(\beta_2, \beta_{12} \mid \beta_0)/2}{MS_E} = \frac{882.7508/2}{0.4059} = 1087.40$$

Reject H_0.

12-91 Consider the X-ray inspection data in Exercise 12-15. Use rads as the response. Build regression models for the data using the following techniques:

(a) All possible regressions.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer? Why?

(a) The min C_p model is: x_1, x_2

$C_p = 3.0$ and $MS_E = 55563.92$

$$\hat{y} = -440.39 + 19.147x_1 + 68.080x_2$$

The min MS_E model is the same as the min C_p.

(b) Same as the model in part (a).
(c) Same as the model in part (a).
(d) Same as the model in part (a).
(e) All methods give the same model with either min C_p or min MS_E.

12-92 Consider the electric power data in Exercise 12-10. Build regression models for the data using the following techniques:

(a) All possible regressions. Find the minimum C_p and minimum MS_E equations.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer?

The default settings for F-to-enter and F-to-remove, equal to 4, in the computer software were used. Different settings can change the models generated by the method.

(a) The min MS_E model is: x_1, x_2, x_3

$C_p = 3.8 \quad MS_E = 134.6$

$$\hat{y} = -162.1 + 0.7487x_1 + 7.691x_2 + 2.343x_3$$

The min C_p model is: x_1, x_2,

$C_p = 3.4 \quad MS_E = 145.7$

$$\hat{y} = 3.92 + 0.5727x_1 + 9.882x_2$$

(b) Same as the min C_p model in part (a)
(c) Same as part min MS_E model in part (a)
(d) Same as part min C_p model in part (a)
(e) The minimum MS_E and forward models all are the same. Stepwise and backward regressions generate the minimum C_p model. The minimum C_p model has fewer regressors and it might be preferred, but MS_E has increased.

12-93 Consider the regression model fit to the coal and limestone mixture data in Exercise 12-17. Use density as the response. Build regression models for the data using the following techniques:

(a) All possible regressions.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer? Why?

(a) The min C_p model is: x_1

$C_p = 1.1$ and $MS_E = 0.0000705$

$$\hat{y} = -0.20052 + 0.467864x_1$$

The min MS_E model is the same as the min C_p.

(b) Same as model in part (a).
(c) Same as model in part (a).
(d) Same as model in part (a).
(e) All methods give the same model with either min C_p or min MS_E.

12-94 Consider the wire bond pull strength data in Exercise 12-12. Build regression models for the data using the following methods:

(a) All possible regressions. Find the minimum C_p and minimum MS_E equations.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer?

The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

(a) The min MS_E model is: x_1, x_3, x_4

$C_p = 2.6 \quad MS_E = 0.6644$

$$\hat{y} = 2.419 + 0.5530x_1 + 0.4790x_3 - 0.12338x_4$$

The min C_p model is: x_3, x_4

$C_p = 1.6 \quad MS_E = 0.7317$

$$\hat{y} = 4.656 + 0.5113x_3 - 0.12418x_4$$

(b) Same as the min C_p model in part (a)
(c) Same as the min C_p model in part (a)
(d) Same as the min C_p model in part (a)
(e) The minimum MS_E and forward models all are the same. Stepwise and backward regressions generate the minimum C_p model. The minimum C_p model has fewer regressors and it might be preferred, but MS_E has increased.

12-95 Consider the gray range modulation data in Exercise 12-19. Use the useful range as the response. Build regression models for the data using the following techniques:

(a) All possible regressions.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer? Why?

(a) The min C_p model is: x_2

$$C_p = 1.2 \text{ and } MS_E = 1178.55$$

$$\hat{y} = 253.06 - 2.5453x_2$$

The min MS_E model is the same as the min C_p.
(b) Same as model in part (a).
(c) Same as model in part (a).
(d) Same as model in part (a).
(e) All methods give the same model with either min C_p or min MS_E.

12-96 Consider the nisin extraction data in Exercise 12-18. Build regression models for the data using the following techniques:

(a) All possible regressions.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer? Why?

(a) The min C_p model is: x_1 ,x_2

$$C_p = 3.0 \text{ and } MS_E = 9.4759$$

$$\hat{y} = -171 + 7.029x_1 + 12.696x_2$$

The min MS_E model is the same as the min C_p.
(b) Same as model in part (a).
(c) Same as model in part (a).
(d) Same as model in part (a).
(e) All methods give the same model with either min C_p or min MS_E.

12-97 Consider the stack loss data in Exercise 12-20. Build regression models for the data using the following techniques:

(a) All possible regressions.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer? Why?
(f) Remove any influential data points and repeat the model building in the previous parts? Does your conclusion in part (e) change?

(a) The min C_p model is: x_1, x_2

$$C_p = 2.9 \text{ and } MS_E = 10.49$$

$$\hat{y} = -50.4 + 0.671x_1 + 1.30x_2$$

The min MS_E model is the same as the min C_p.
(b) Same as model in part (a).
(c) Same as model in part (a).
(d) Same as model in part (a).
(e) All methods give the same model with either min C_p or min MS_E.
(f) There are no observations with a Cook's distance greater than 1 so the results will be the same.

12-98 Consider the NHL data in Exercise 12-22. Build regression models for these data with regressors *GF* through *FG* using the following methods:

(a) All possible regressions. Find the minimum C_p and minimum MS_E equations.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Which model would you prefer?

The default settings for F-to-enter and F-to-remove for Minitab were used. Different settings can change the models generated by the method.

(a)

Best Subsets Regression: W versus GF, GA, ...

```
Response is W

                                                            P
                                              P P         P K S S
                                            A P C P B A S P P H H
                       Mallows          G G D G T E M V H G C G G F
Vars  R-Sq  R-Sq(adj)      C-p       S  F A V F G N I G T A T F A G
   1  52.8       51.1     74.3  5.0233  X
   1  49.3       47.5     81.6  5.2043                            X
   2  86.5       85.5      4.7  2.7378  X X
   2  79.3       77.8     19.9  3.3832  X                 X
   3  87.7       86.2      4.2  2.6648  X X                     X
   3  87.3       85.8      5.0  2.7045  X X             X
   4  88.7       86.9      4.0  2.6011  X X             X       X
   4  88.2       86.3      5.0  2.6561  X X           X         X
   5  89.5       87.3      4.4  2.5620  X X             X X X
   5  89.3       87.0      4.8  2.5866  X X               X X   X
   6  91.3       89.1      2.4  2.3728  X X             X X X   X
   6  89.9       87.3      5.4  2.5621  X X X             X X   X
   7  92.3       89.8      2.4  2.2894  X X X           X X X   X
   7  91.6       89.0      3.8  2.3874  X X       X     X X X   X
   8  92.6       89.8      3.7  2.2954  X X X     X     X X X   X
   8  92.6       89.8      3.7  2.2967  X X X         X X X X   X
   9  92.7       89.5      5.4  2.3295  X X X     X     X X X X X
   9  92.7       89.5      5.4  2.3309  X X X         X X X X X X
  10  92.8       89.0      7.3  2.3829  X X X X   X     X X X X X
  10  92.8       89.0      7.3  2.3833  X X X   X X     X X X X X
  11  92.8       88.5      9.2  2.4402  X X X X   X   X X X X X X
  11  92.8       88.5      9.2  2.4406  X X X   X X   X X X X X X
  12  92.9       87.9     11.0  2.4936  X X X X   X X X X X X X X
  12  92.9       87.9     11.0  2.4939  X X X   X X X X X X X X X
  13  92.9       87.2     13.0  2.5702  X X X X X X X X X X X X X
  13  92.9       87.2     13.0  2.5703  X X X X   X X X X X X X X X
  14  92.9       86.3     15.0  2.6544  X X X X X X X X X X X X X X
```

From the output the minimum CP model and minimum MSE model are the same.
The regressors are GF, GA, ADV, SHT, PPGA, PKPCT, SHGA. The computer output for this model follows.

Regression Analysis: W versus GF, GA, ADV, SHT, PPGA, PKPCT, SHGA

```
The regression equation is
W = 457 + 0.182 GF - 0.187 GA - 0.0375 ADV + 0.256 SHT - 1.44 PPGA - 4.94 PKPCT
    + 0.489 SHGA

Predictor       Coef  SE Coef      T      P
Constant       457.3    138.5   3.30  0.003
```

```
GF           0.18233  0.02018   9.04  0.000
GA          -0.18657  0.03317  -5.62  0.000
ADV         -0.03753  0.02282  -1.65  0.114
SHT          0.25638  0.09826   2.61  0.016
PPGA         -1.4420   0.4986  -2.89  0.008
PKPCT         -4.935    1.679  -2.94  0.008
SHGA          0.4893   0.1785   2.74  0.012

S = 2.28935   R-Sq = 92.3%   R-Sq(adj) = 89.8%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       7  1380.70  197.24  37.63  0.000
Residual Error  22   115.30    5.24
Total           29  1496.00
```

The model is

$$\hat{y} = 457 + 0.182GF - 0.187GA - 0.0375ADV + 0.256SHT - 1.44PPGA - 4.94PKPCT + 0.489SHGA$$

(b) **Stepwise Regression: W versus GF, GA, ...**

```
  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

Response is W on 14 predictors, with N = 30

Step              1       2       3       4
Constant     -8.574  40.271  38.311  43.164

GF            0.212   0.182   0.182   0.187
T-Value        5.59    8.68    8.92    9.26
P-Value       0.000   0.000   0.000   0.000

GA                   -0.179  -0.179  -0.167
T-Value               -8.20   -8.40   -7.51
P-Value               0.000   0.000   0.000

SHGA                           0.27    0.29
T-Value                        1.58    1.76
P-Value                       0.126   0.090

SHT                                  -0.026
T-Value                               -1.51
P-Value                               0.143

S              5.02    2.74    2.66    2.60
R-Sq          52.77   86.47   87.66   88.69
R-Sq(adj)     51.08   85.47   86.23   86.88
Mallows C-p    74.3     4.7     4.2     4.0
```

The selected model from Stepwise Regression has four regressors GF, GA, SHT, SHGA. The computer output for this model follows.

Regression Analysis: W versus GF, GA, SHT, SHGA

```
The regression equation is
W = 43.2 + 0.187 GF - 0.167 GA - 0.0259 SHT + 0.293 SHGA

Predictor      Coef  SE Coef      T      P
Constant     43.164    8.066   5.35  0.000
GF          0.18677  0.02016   9.26  0.000
```

```
GA         -0.16683  0.02221  -7.51  0.000
SHT        -0.02587  0.01710  -1.51  0.143
SHGA         0.2926   0.1660   1.76  0.090

S = 2.60115   R-Sq = 88.7%   R-Sq(adj) = 86.9%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       4  1326.85  331.71  49.03  0.000
Residual Error  25   169.15    6.77
Total           29  1496.00
```

The model is

$$\hat{y} = 43.2 + 0.187GF - 0.167GA - 0.0259SHT + 0.293SHGA$$

(c) **Stepwise Regression: W versus GF, GA, ...**

```
Forward selection.  Alpha-to-Enter: 0.25

Response is W on 14 predictors, with N = 30

Step              1       2       3       4
Constant     -8.574  40.271  38.311  43.164

GF            0.212   0.182   0.182   0.187
T-Value        5.59    8.68    8.92    9.26
P-Value       0.000   0.000   0.000   0.000

GA                   -0.179  -0.179  -0.167
T-Value               -8.20   -8.40   -7.51
P-Value               0.000   0.000   0.000

SHGA                           0.27    0.29
T-Value                        1.58    1.76
P-Value                       0.126   0.090

SHT                                  -0.026
T-Value                               -1.51
P-Value                               0.143

S              5.02    2.74    2.66    2.60
R-Sq          52.77   86.47   87.66   88.69
R-Sq(adj)     51.08   85.47   86.23   86.88
Mallows C-p    74.3     4.7     4.2     4.0
```

The model selected by Forward Selection is the same as part (b).

(d) **Stepwise Regression: W versus GF, GA, ...**

```
Backward elimination.  Alpha-to-Remove: 0.1

Response is W on 14 predictors, with N = 30

Step               1        2        3        4        5        6
Constant       512.2    513.0    511.3    524.0    520.5    507.3

GF             0.164    0.164    0.164    0.166    0.167    0.173
T-Value         4.46     4.97     5.24     5.53     5.66     7.65
P-Value        0.000    0.000    0.000    0.000    0.000    0.000

GA            -0.183   -0.184   -0.184   -0.186   -0.190   -0.191
```

```
T-Value         -3.83    -4.58    -4.74    -4.95    -5.43    -5.62
P-Value         0.002    0.000    0.000    0.000    0.000    0.000

ADV            -0.054   -0.054   -0.046   -0.043   -0.040   -0.036
T-Value         -0.25    -0.25    -1.52    -1.48    -1.48    -1.52
P-Value         0.808    0.802    0.147    0.157    0.156    0.145

PPGF            0.089    0.087    0.047    0.031    0.022
T-Value          0.08     0.08     0.59     0.43     0.34
P-Value         0.938    0.937    0.565    0.671    0.739

PCTG             -0.1     -0.1
T-Value         -0.04    -0.04
P-Value         0.971    0.971

PEN           -0.1632  -0.1628  -0.1646  -0.0365  -0.0039  -0.0043
T-Value         -0.54    -0.56    -0.59    -0.38    -0.86    -1.00
P-Value         0.598    0.586    0.564    0.706    0.400    0.330

BMI             -0.13    -0.13    -0.13
T-Value         -0.45    -0.47    -0.49
P-Value         0.658    0.647    0.632

AVG              13.1     13.1     13.2      2.7
T-Value          0.53     0.54     0.57     0.34
P-Value         0.606    0.594    0.574    0.735

SHT              0.29     0.29     0.29     0.31     0.31     0.30
T-Value          2.19     2.30     2.40     2.65     2.71     2.76
P-Value         0.045    0.035    0.028    0.016    0.014    0.012

PPGA            -1.60    -1.60    -1.61    -1.66    -1.64    -1.60
T-Value         -2.50    -2.61    -2.71    -2.90    -2.96    -3.02
P-Value         0.025    0.019    0.015    0.009    0.008    0.007

PKPCT            -5.5     -5.6     -5.6     -5.7     -5.7     -5.5
T-Value         -2.54    -2.66    -2.77    -2.96    -3.02    -3.09
P-Value         0.023    0.017    0.013    0.008    0.007    0.006

SHGF             0.11     0.11     0.11     0.09     0.09     0.10
T-Value          0.54     0.56     0.62     0.54     0.59     0.62
P-Value         0.600    0.584    0.541    0.593    0.565    0.540

SHGA             0.61     0.61     0.61     0.57     0.54     0.53
T-Value          2.34     2.42     2.59     2.64     2.78     2.85
P-Value         0.033    0.028    0.019    0.016    0.012    0.010

FG               0.00
T-Value          0.02
P-Value         0.981

S                2.65     2.57     2.49     2.44     2.38     2.33
R-Sq            92.94    92.93    92.93    92.84    92.79    92.75
R-Sq(adj)       86.34    87.19    87.95    88.46    88.99    89.48
Mallows C-p      15.0     13.0     11.0      9.2      7.3      5.4


Step                7        8        9
Constant        496.5    457.3    417.5

GF              0.178    0.182    0.177
T-Value          8.53     9.04     8.57
P-Value         0.000    0.000    0.000
```

```
GA             -0.189  -0.187  -0.187
T-Value         -5.66   -5.62   -5.43
P-Value         0.000   0.000   0.000

ADV            -0.038  -0.038
T-Value         -1.67   -1.65
P-Value         0.109   0.114

PPGF
T-Value
P-Value

PCTG
T-Value
P-Value

PEN           -0.0039
T-Value         -0.94
P-Value         0.358

BMI
T-Value
P-Value

AVG
T-Value
P-Value

SHT             0.298   0.256   0.238
T-Value          2.76    2.61    2.35
P-Value         0.012   0.016   0.028

PPGA            -1.58   -1.44   -1.34
T-Value         -3.03   -2.89   -2.62
P-Value         0.006   0.008   0.015

PKPCT            -5.4    -4.9    -4.6
T-Value         -3.08   -2.94   -2.65
P-Value         0.006   0.008   0.014

SHGF
T-Value
P-Value

SHGA             0.51    0.49    0.39
T-Value          2.83    2.74    2.23
P-Value         0.010   0.012   0.036

FG
T-Value
P-Value

S                2.30    2.29    2.37
R-Sq            92.60   92.29   91.34
R-Sq(adj)       89.79   89.84   89.09
Mallows C-p       3.7     2.4     2.4
```

The model selected by Backward Selection includes GF, GA, SHT, PPGA, PKPCT, SHGA. The computer output for this model follows.

Regression Analysis: W versus GF, GA, SHT, PPGA, PKPCT, SHGA

```
The regression equation is
W = 418 + 0.177 GF - 0.187 GA + 0.238 SHT - 1.34 PPGA - 4.58 PKPCT + 0.387 SHGA

Predictor       Coef  SE Coef      T      P
Constant       417.5    141.3   2.95  0.007
GF           0.17679  0.02062   8.57  0.000
GA          -0.18677  0.03438  -5.43  0.000
SHT           0.2377   0.1012   2.35  0.028
PPGA         -1.3426   0.5130  -2.62  0.015
PKPCT         -4.578    1.725  -2.65  0.014
SHGA          0.3869   0.1734   2.23  0.036

S = 2.37276   R-Sq = 91.3%   R-Sq(adj) = 89.1%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       6  1366.51  227.75  40.45  0.000
Residual Error  23   129.49    5.63
Total           29  1496.00
```

The model is

$$\hat{y} = 418 + 0.177GF - 0.187GA + 0.238SHT - 1.34PPGA - 4.58PKPCT + 0.387SHGA$$

(e) There are several reasonable choices.
The seven-variable model GF, GA, ADV, SHT, PPGA, PKPCT, SHGA with minimum C_p is a good choice. It has MSE not much larger than the MSE in the full model.

The four-variable model GF, GA, SHT, SHGA from Stepwise Regression(and Forward Selection) is a simpler model with Cp = 4.0 < p =5 and good R-squared.

Even the three-variable model GF, GA, SHGA is reasonable. It is still simpler with a good R-squared. The Cp = 4.2 and this is only slightly greater than p = 4. However, the MSE for this model is somewhat higher than for the six-variable model.

12-99 Use the football data in Exercise 12-21 to build regression models using the following techniques:

(a) All possible regressions. Find the equations that minimize MS_E and that minimize C_p.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the various models obtained. Which model seems "best,'' and why?

(a) The computer output follows. The first model in the table with seven variables minimizes MS_E and Cp.

Best Subsets Regression: Pts versus Att, Comp, ...

```
PctComp, YdsPerAtt, PctTD, PctInt

Response is Pts

                                                    Y
                                                    d
                                                P   s
                                                c   p         P
                                                t   e   P     c
                                              C C   r   c     t
                                            A o o Y A   t L I I
                                            t m m d t T T n n n
Vars   R-Sq  R-Sq(adj)  Mallows C-p       S t p p s t D D g t t
```

1	67.2	66.1	38635.3	5.2187					X					
1	64.8	63.7	41381.3	5.4009			X							
2	85.1	84.1	17511.0	3.5748			X				X			
2	83.9	82.8	18943.5	3.7180					X					X
3	95.3	94.8	5461.8	2.0348			X				X			X
3	93.1	92.4	8064.5	2.4708			X			X			X	
4	100.0	100.0	5.3	0.14621			X		X		X			X
4	98.5	98.3	1730.7	1.1713			X		X	X			X	
5	100.0	100.0	6.1	0.14558			X	X	X		X			X
5	100.0	100.0	6.2	0.14600		X	X		X		X			X
6	100.0	100.0	5.2	0.13990	X		X	X	X		X			X
6	100.0	100.0	6.4	0.14362	X	X	X		X		X			X
7	100.0	100.0	5.1	0.13632	X		X	X	X	X	X			X
7	100.0	100.0	7.1	0.14277	X		X	X	X		X		X	X
8	100.0	100.0	7.0	0.13892	X		X	X	X	X	X		X	X
8	100.0	100.0	7.1	0.13924	X	X	X	X	X	X	X			X
9	100.0	100.0	9.0	0.14203	X		X	X	X	X	X	X	X	X
9	100.0	100.0	9.0	0.14204	X	X	X	X	X	X	X		X	X
10	100.0	100.0	11.0	0.14537	X	X	X	X	X	X	X	X	X	X

The computer output for this model follows.

Regression Analysis: RatingPts versus Att, PctComp, ...

```
The regression equation is
RatingPts = - 0.69 + 0.00738 Att + 0.827 PctComp - 0.00150 Yds + 4.82 YdsPerAtt
            + 0.0702 TD + 3.04 PctTD - 4.19 PctInt

Predictor         Coef    SE Coef        T      P
Constant        -0.693      1.264    -0.55  0.589
Att           0.007381   0.003258     2.27  0.033
PctComp       0.826817   0.008980    92.07  0.000
Yds         -0.0015027  0.0006166    -2.44  0.023
YdsPerAtt       4.8206     0.2555    18.87  0.000
TD             0.07025    0.04601     1.53  0.140
PctTD           3.0386     0.2013    15.10  0.000
PctInt        -4.19493    0.03545  -118.34  0.000

S = 0.136317   R-Sq = 100.0%   R-Sq(adj) = 100.0%

Analysis of Variance

Source          DF       SS      MS         F      P
Regression       7  2489.08  355.58  19135.65  0.000
Residual Error  24     0.45    0.02
Total           31  2489.52
```

(b) Stepwise regression selects the four-variable model `YdsPerAtt`, `PctInt`, `PctTD`, `PctComp`.

Stepwise Regression: Pts versus Att, Comp, ...

```
  Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

Response is Pts on 10 predictors, with N = 32

Step              1        2       3        4
Constant     14.195   24.924  32.751    2.128

YdsperAtt    10.092   10.296   7.432    4.214
T-Value        7.84    11.21    8.78    73.46
P-Value       0.000    0.000   0.000    0.000
```

```
PctInt                 -4.786  -4.932   -4.164
T-Value                 -5.49   -7.90  -118.11
P-Value                 0.000   0.000    0.000

PctTD                           3.187    3.310
T-Value                          5.36   101.17
P-Value                         0.000    0.000

PctComp                                 0.8284
T-Value                                  96.04
P-Value                                  0.000

S               5.22     3.72    2.66    0.146
R-Sq           67.18    83.90   92.06    99.98
R-Sq(adj)      66.09    82.79   91.21    99.97
Mallows C-p  38635.3  18943.5  9332.6      5.3
```

The computer output for this model follows.

Regression Analysis: RatingPts versus YdsPerAtt, PctInt, PctTD, PctComp

```
The regression equation is
RatingPts = 2.13 + 4.21 YdsPerAtt - 4.16 PctInt + 3.31 PctTD + 0.828 PctComp

Predictor      Coef   SE Coef        T      P
Constant     2.1277    0.4224     5.04  0.000
YdsPerAtt   4.21364   0.05736    73.46  0.000
PctInt     -4.16391   0.03526  -118.11  0.000
PctTD       3.31029   0.03272   101.17  0.000
PctComp    0.828385  0.008626    96.04  0.000

S = 0.146206   R-Sq = 100.0%   R-Sq(adj) = 100.0%

Analysis of Variance

Source          DF       SS      MS         F      P
Regression       4  2488.95  622.24  29108.69  0.000
Residual Error  27     0.58    0.02
Total           31  2489.52

The model is
```

$$\hat{y} = 2.13 + 4.21 YdsPerAtt - 4.16 PctInt + 3.31 PctTD + 0.828 PctComp$$

(c) Forward selection shown below selects the same four-variable model `YdsPerAtt,PctInt,PctTD,PctComp` as in part (b).

Stepwise Regression: Pts versus Att, Comp, ...

```
Forward selection.  Alpha-to-Enter: 0.25

Response is Pts on 10 predictors, with N = 32

Step              1        2       3        4
Constant     14.195   24.924  32.751    2.128

YdsperAtt    10.092   10.296   7.432    4.214
T-Value        7.84    11.21    8.78    73.46
P-Value       0.000    0.000   0.000    0.000

PctInt                -4.786  -4.932   -4.164
```

```
T-Value                -5.49   -7.90  -118.11
P-Value                0.000   0.000    0.000

PctTD                          3.187    3.310
T-Value                         5.36   101.17
P-Value                        0.000    0.000

PctComp                                0.8284
T-Value                                 96.04
P-Value                                 0.000

S                5.22     3.72    2.66    0.146
R-Sq            67.18    83.90   92.06    99.98
R-Sq(adj)       66.09    82.79   91.21    99.97
Mallows C-p  38635.3  18943.5  9332.6      5.3
```

(d) Backward elimination shown below selects the six-variable model
`Att, PctComp, Yds, YdsPerAtt, PctTD, PctInt.`
It is similar to the model with minimum MS_E except variable `TD` is excluded.

Stepwise Regression: Pts versus Att, Comp, ...

```
Backward elimination.  Alpha-to-Remove: 0.1


Response is Pts on 10 predictors, with N = 32


Step                1         2         3         4         5
Constant      -0.6871   -0.7140   -0.6751   -0.6928   -0.1704

Att            0.0074    0.0074    0.0074    0.0074    0.0055
T-Value          1.56      2.02      2.22      2.27      1.78
P-Value         0.133     0.056     0.036     0.033     0.088

Comp            0.000
T-Value          0.02
P-Value         0.982

PctComp        0.8253    0.8264    0.8264    0.8268    0.8289
T-Value         16.92     86.99     89.30     92.07     91.04
P-Value         0.000     0.000     0.000     0.000     0.000

Yds          -0.00159  -0.00158  -0.00156  -0.00150  -0.00083
T-Value         -1.65     -2.20     -2.39     -2.44     -1.88
P-Value         0.115     0.038     0.026     0.023     0.072

YdsperAtt        4.86      4.85      4.85      4.82      4.55
T-Value         11.78     16.28     17.69     18.87     24.46
P-Value         0.000     0.000     0.000     0.000     0.000

TD              0.075     0.075     0.074     0.070
T-Value          1.46      1.50      1.53      1.53
P-Value         0.158     0.148     0.139     0.140

PctTD           3.018     3.019     3.019     3.039     3.341
T-Value         13.47     13.82     14.15     15.10     96.41
P-Value         0.000     0.000     0.000     0.000     0.000

Lng            0.0001    0.0001
T-Value          0.05      0.05
```

```
P-Value         0.960     0.961

Int             0.011     0.011     0.010
T-Value          0.31      0.33      0.33
P-Value         0.756     0.747     0.743

PctInt         -4.241    -4.240    -4.238    -4.195    -4.177
T-Value        -27.73    -29.74    -31.63   -118.34   -121.82
P-Value         0.000     0.000     0.000     0.000     0.000

S               0.145     0.142     0.139     0.136     0.140
R-Sq            99.98     99.98     99.98     99.98     99.98
R-Sq(adj)       99.97     99.97     99.98     99.98     99.98
Mallows C-p      11.0       9.0       7.0       5.1       5.2
```

The computer output for this model follows.

Regression Analysis: RatingPts versus Att, PctComp, ...

```
The regression equation is
RatingPts = - 0.17 + 0.00550 Att + 0.829 PctComp - 0.000826 Yds + 4.55
YdsPerAtt
            + 3.34 PctTD - 4.18 PctInt

Predictor        Coef    SE Coef        T      P
Constant       -0.170      1.248    -0.14  0.893
Att          0.005499   0.003095     1.78  0.088
PctComp      0.828933   0.009105    91.04  0.000
Yds        -0.0008259  0.0004398    -1.88  0.072
YdsPerAtt      4.5455     0.1858    24.46  0.000
PctTD         3.34144    0.03466    96.41  0.000
PctInt       -4.17685    0.03429  -121.82  0.000

S = 0.139897   R-Sq = 100.0%   R-Sq(adj) = 100.0%

Analysis of Variance

Source          DF       SS      MS         F      P
Regression       6  2489.03  414.84  21196.51  0.000
Residual Error  25     0.49    0.02
Total           31  2489.52
```

The model is

$$\hat{y} = -0.17 + 0.0055Att + 0.8289PctComp - 0.0008Yds + 4.5455YdsPerAtt + 3.3414PctTd - 4.1769PctInt$$

(e) The four variable model (`PctComp, YdsPerAtt, PctTD, PctInt`) has the second minimum Cp and also has small MS_E and large adjusted R-squared. It is also a model with the few regressors so it is preferred.

12-100 Consider the arsenic data in Exercise 12-16. Use arsenic in nails as the response and age, drink use, and cook use as the regressors. Build regression models for the data using the following techniques:

(a) All possible regressions.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer? Why?
(f) Now construct an indicator variable and add the person's sex to the list of regressors. Repeat the model building in the previous parts. Does your conclusion in part (e) change?

(a) The min C_p model is: x_1

$$C_p = 0.0 \text{ and } MS_E = 0.2298$$

$$\hat{y} = -0.038 + 0.00850x_1$$

The min MS_E model is the same as the min C_p model

(b) The full model that contains all 3 variables

$$\hat{y} = 0.001 - 0.00858x_1 - 0.021x_2 - 0.010x_3$$

where $x_1 = AGE \quad x_2 = DrinkUse \quad x_3 = CookUse$

(c) No variables are selected

(d) The min C_p model has only the intercept term with $C_p = -0.5$ and $MS_E = 0.2372$

The min MS_E model is the same as the min C_p in part (a).

(e) None of the variables seem to be good predictors of arsenic in nails based on the models above (none of the variables are significant).

12-101 Consider the gas mileage data in Exercise 12-11. Build regression models for the data from the numerical regressors using the following techniques:

(a) All possible regressions.
(b) Stepwise regression.
(c) Forward selection.
(d) Backward elimination.
(e) Comment on the models obtained. Which model would you prefer? Why?
(f) Now construct indicator variable for *trns* and *drv* and add these to the list of regressors. Repeat the model building in the previous parts. Does your conclusion in part (e) change?

This analysis includes the emissions variables hc, co, and co2. It would be reasonable to consider models without these variables as regressors.

Best Subsets Regression: mpg versus cid, rhp, ...

Response is mpg

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	cid	rhp	etw	cmp	axle	n/v	hc	co	co2
1	66.0	64.2	26.5	3.4137			X						
1	59.1	57.0	35.3	3.7433	X								
2	81.6	79.5	8.6	2.5813	X		X						
2	78.1	75.7	13.0	2.8151			X		X				
3	88.8	86.8	1.3	2.0718	X		X		X				
3	88.8	86.8	1.4	2.0755			X		X	X			
4	90.3	87.8	1.5	1.9917	X		X		X				X
4	89.9	87.3	2.0	2.0302			X		X	X		X	
5	90.7	87.6	2.9	2.0057	X		X	X	X				X
5	90.7	87.6	2.9	2.0064	X		X		X		X		X
6	91.0	87.2	4.5	2.0442	X		X		X		X	X	X
6	91.0	87.1	4.5	2.0487	X		X	X	X		X		X
7	91.3	86.6	6.2	2.0927		X	X		X	X	X	X	X
7	91.2	86.4	6.3	2.1039	X		X	X	X		X	X	X
8	91.4	85.6	8.0	2.1651		X	X	X	X	X	X	X	X
8	91.4	85.6	8.1	2.1654	X	X	X	X	X		X	X	X
9	91.4	84.4	10.0	2.2562	X	X	X	X	X	X	X	X	X

(a) The minimum C_p (1.3) model is:

$\hat{y} = 61.001 - 0.02076x_{cid} - 0.00354x_{etw} - 3.457x_{axle}$

The minimum MSE (4.0228) model is:

$\hat{y} = 49.5 - 0.017547x_{cid} - 0.0034252x_{etw} + 1.29x_{cmp} - 3.184x_{axle} - 0.0096x_{c02}$

(b) $\hat{y} = 63.31 - 0.0178x_{cid} - 0.00375x_{etw} - 3.3x_{axle} - 0.0084x_{c02}$

(c) Same model as the min MS_E equation in part (a)

(d) $\hat{y} = 45.18 - 0.00321x_{etw} - 4.4x_{axle} + 0.385x_{n/v}$

(e) The minimum C_p model is preferred because it has a very low MSE as well (4.29)

(f) Only one indicator variable is used for transmission to distinguish the automatic from manual types and two indicator variables are used for drv:

x_{trans} = 0 for automatic (L4, L5) and 1 for manual (M6) and

x_{drv1} = 0 if drv = 4 or R and 1 if drv =F; x_{drv2} = 0 if drv = 4 or F and 1 if drv = R.

The minimum C_p (4.0) model is the same as the minimum MSE (2.267) model:

$$\hat{y} = 10 - 0.0038023x_{etw} + 3.936x_{cmp} + 15.216x_{co} - 0.011118x_{c02} - 7.401x_{trans} +$$
$$3.6131x_{drv1} + 2.342x_{drv2}$$

Stepwise:

$\hat{y} = 39.12 - 0.0044x_{etw} + 0.271x_{n/v} - 4.5x_{trns} + 3.2x_{drv1} + 1.7x_{drv2}$

Forward selection:

$\hat{y} = 41.12 - 0.00377x_{etw} + 0.336x_{n/v} - 2.1x_{axle} - 3.4x_{trans} +$
$2.1x_{drv1} + 2x_{drv2}$

Backward selection: same as minimum C_p and minimum MSE.
Prefer the model giving the minimum C_p and minimum MSE.

12-102 When fitting polynomial regression models, we often subtract $\bar{x}$ from each $\bar{x}$ value to produce a "centered'' regressor $x' = x - \bar{x}$. This reduces the effects of dependencies among the model terms and often leads to more accurate estimates of the regression coefficients. Using the data from Exercise 12-84, fit the model $Y = \beta_0^* + \beta_1^* x' + \beta_{11}^*(x')^2 + \in$.

(a) Use the results to estimate the coefficients in the uncentered model $Y = \beta_0 + \beta_1 x + \beta_{11}x^2 + \in$. Predict *y* when *x* = 285°F. Suppose that you use a standardized variable $x' = (x - \bar{x})/s_x$ where s_x is the standard deviation of *x* in constructing a polynomial regression model. Fit the model $Y = \beta_0^* + \beta_1^* x' + \beta_{11}^*(x')^2 + \in$.

(b) What value of *y* do you predict when x = 285°F?

(c) Estimate the regression coefficients in the unstandardized model $Y = \beta_0 + \beta_1 x + \beta_{11}x^2 + \in$.

(d) What can you say about the relationship between SS_E and R^2 for the standardized and unstandardized models?

(e) Suppose that $y' = (y - \bar{y})/s_y$ is used in the model along with x'. Fit the model and comment on the relationship between SS_E and R^2 in the standardized model and the unstandardized model.

$$\hat{y} = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2$$
$$\hat{y} = 759.395 - 7.607x' - 0.331(x')^2$$
$$\hat{y} = 759.395 - 7.607(x - 297.125) - 0.331(x - 297.125)^2$$
$$\hat{y} = -26202.14 + 189.09x - 0.331x^2$$

(a) $\hat{y} = 759.395 - 90.783x' - 47.166(x')^2$, where $x' = \frac{x - \bar{x}}{S_x}$

(b) At x = 285 $x' = \frac{285 - 297.125}{11.9336} = -1.016$

$$\hat{y} = 759.395 - 90.783(-1.106) - 47.166(-1.106)^2 = 802.943 \text{ psi}$$

(c) $\hat{y} = 759.395 - 90.783\left(\frac{x - 297.125}{11.9336}\right) - 47.166\left(\frac{x - 297.125}{11.9336}\right)^2$

$$\hat{y} = 759.395 - 7.607(x - 297.125) - 0.331(x - 297.125)^2$$
$$\hat{y} = -26204.14 + 189.09x - 0.331x^2$$

(d) They are the same.

(e) $\hat{y}' = 0.385 - 0.847x' - 0.440(x')^2$

where $y' = \frac{y - \bar{y}}{S_y}$ and $x' = \frac{x - \bar{x}}{S_x}$

The proportion of total variability explained is the same for both the standardized and un-standardized models. Therefore, R^2 is the same for both models.

$y' = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2$ where $y' = \frac{y - \bar{y}}{S_y}$ and $x' = \frac{x - \bar{x}}{S_x}$

$$y' = \beta_0^* + \beta_1^* x' + \beta_{11}^* (x')^2$$

12-103 Consider the data in Exercise 12-87. Use all the terms in the full quadratic model as the candidate regressors.

(a) Use forward selection to identify a model.
(b) Use backward elimination to identify a model.
(c) Compare the two models obtained in parts (a) and (b). Which model would you prefer and why?

The default settings for F-to-enter and F-to-remove, equal to 4, were used. Different settings can change the models generated by the method.

(a) $\hat{y} = -0.304 + 0.083x_1 - 0.031x_3 + 0.004x_2^2$

$$C_p = 4.04 \quad MS_E = 0.004$$

(b) $\hat{y} = -0.256 + 0.078x_1 + 0.022x_2 - 0.042x_3 + 0.0008x_3^2$

$$C_p = 4.66 \quad MS_E = 0.004$$

(c) The forward selection model in part (a) is more parsimonious with a lower C_p and equivalent MS_E. Therefore, we prefer the model in part (a).

12-104 We have used a sample of 30 observations to fit a regression model. The full model has nine regressors, the variance estimate is $\hat{\sigma}^2 = MS_E = 100$, and $R^2 = 0.92$.

(a) Calculate the *F*-statistic for testing significance of regression. Using α = 0.05, what would you conclude?
(b) Suppose that we fit another model using only four of the original regressors and that the error sum of squares for

this new model is 2200. Find the estimate of σ^2 for this new reduced model. Would you conclude that the reduced model is superior to the old one? Why?

(c) Find the value of C_p for the reduced model in part (b). Would you conclude that the reduced model is better than the old model?

n = 30, k = 9, p =9 + 1 = 10 in full model.

(a) $\hat{\sigma}^2 = MS_E = 100$ $\quad R^2 = 0.92$

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}}$$

$$SS_E = MS_E(n-p) = 100(30-10) = 2000$$

$$0.92 = 1 - \frac{2000}{S_{yy}}$$

$$25000 = S_{yy}$$

$$SS_R = S_{yy} - SS_E = 25000 - 2000 = 23000$$

$$MS_R = \frac{SS_R}{k} = \frac{23000}{9} = 2555.56$$

$$f_0 = \frac{MS_R}{MS_E} = \frac{2555.56}{100} = 25.56$$

$$f_{.05,9,20} = 2.39$$

$$f_0 > f_{\alpha,9,20}$$

Reject H_0 and conclude at least one β_j is significant at $\alpha = 0.05$.

(b) k = 4 p = 5 $\quad SS_E = 2200$

$$MS_E = \frac{SS_E}{n-p} = \frac{2200}{30-5} = 88$$

Yes, MS_E is reduced with new model (k = 4).

(c) $C_p = \dfrac{SS_E(p)}{\hat{\sigma}^2} - n + 2p \quad C_p = \dfrac{2200}{100} - 30 + 2(5) = 2$

Yes, C_p is reduced from the full model.

12-105 A sample of 25 observations is used to fit a regression model in seven variables. The estimate of σ^2 for this full model is $MS_E = 10$.

(a) A forward selection algorithm has put three of the original seven regressors in the model. The error sum of squares for the three-variable model is $SS_E = 300$. Based on C_p, would you conclude that the three-variable model has any remaining bias?

(b) After looking at the forward selection model in part (a), suppose you could add one more regressor to the model. This regressor will reduce the error sum of squares to 275. Will the addition of this variable improve the model? Why?

n = 25 $\quad$ k = 7 $\quad$ p = 8 $\quad MS_{E(\text{full})} = 10$

(a) p =4 $\quad SS_E = 300$

$$MS_E = \frac{SS_E}{n-p} = \frac{300}{25-4} = 14.29$$

$$C_p = \frac{SS_E}{MS_{E(full)}} - n + 2p$$
$$= \frac{300}{10} - 25 + 2(4)$$
$$= 5 + 8 = 13$$

YES, C_p > p

(b) p = 5 $SS_E = 275$

$$MS_E = \frac{SS_E}{n-p} = \frac{275}{30-5} = 11 \quad C_p = \frac{275}{10} - 25 + 2(5) = 12.5$$

Yes, both MS_E and C_p are reduced.

Supplemental Exercises

12-106 Consider the following computer output.

The regression equation is
Y = 517 + 11.5 x1 – 8.14 x2 + 10.9 x3

Predictor	Coef	SE Coef	T	P
Constant	517.46	11.76	?	?
x1	11.4720	?	36.50	?
x2	–8.1378	0.1969	?	?
x3	10.8565	0.6652	?	?

S = 10.2560 R-Sq = ? R-Sq (adj) = ?

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	?	347300	115767	?	?
Residual error	16	?	105		
Total	19	348983			

(a) Fill in the missing values. Use bounds for the *P*-values.
(b) Is the overall model significant at $\alpha = 0.05$? Is it significant at $\alpha = 0.01$?
(c) Discuss the contribution of the individual regressors to the model.

(a) The missing quantities are as follows:

$T_{Constant} = \dfrac{Coef}{SE\ Coef} = \dfrac{517.46}{11.76} = 44.0017$

From the t table with 16 degrees of freedom, P-value$_{Constant}$ < 2(0.0005), so P-value$_{Constant}$ < 0.001

$T_{x1} = \dfrac{Coef}{SE\ Coef}$, SE Coef$_{x1}$ $= \dfrac{Coef}{T_{x1}} = \dfrac{11.4720}{36.50} = 0.3143$

P-valule$_{x1}$ < 2(0.0005), so P-valule$_{x1}$ < 0.001

$T_{X2} = \dfrac{Coef}{SE\ Coef} = \dfrac{-8.1378}{0.1969} = -41.3296$

P-valule$_{x2}$ < 2(0.0005), so P-valule$_{x2}$ < 0.001

$T_{x3} = \dfrac{Coef}{SE\ Coef} = \dfrac{10.8565}{0.6652} = 16.3207$

P-valule$_{x3}$ < 2(0.0005), so P-valule$_{x3}$ < 0.001

Regression DF = 19 – 16 = 3

$SS_{Error} = SS_{Total} - SS_{Regression} = 348983 - 347300 = 1683$

$F = \dfrac{MS_{Regression}}{MS_{Error}} = \dfrac{115767}{105} = 1102.543$

P-value < 0.01.

R-Squared = 347300/348983 = 0.995
R-Squared Adjusted = 1 – (1683/16)/(348943/19) = 0.994

(b) Because the P-value from the F-test is less than α = 0.05 and less than α = 0.01, we reject the H_0 for either α value and conclude that at least one regressor significantly contributes to the model.

(c) Because the P-value from the t-test for the x1, x2, and x3 variables are less than α = 0.05, we reject the H_0's and conclude that each individual regressor contributes significantly to the model.

12-107 Consider the following inverse of the model matrix:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.893758 & -0.028245 & -0.0175641 \\ -0.028245 & 0.0013329 & 0.0001547 \\ -0.017564 & 0.0001547 & 0.0009108 \end{bmatrix}$$

(a) How many variables are in the regression model?
(b) If the estimate of σ^2 is 50, what is the estimate of the variance of each regression coefficient?
(c) What is the standard error of the intercept?

(a) Because the matrix is 3 x 3 two regressors are in the regression model. The intercept is also in the model.

(b) $\text{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 C$

Therefore, the variances of the two variables regression coefficients are: $50(0.0013329) = 0.066645$ and $50(0.0009108) = 0.04554$

(c) $se(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 C_{11}} = 6.6849$

12-108 The data shown in Table E12-14 represent the thrust of a jet-turbine engine (y) and six candidate regressors: x_1 = primary speed of rotation, x_2 = secondary speed of rotation, x_3 = fuel flow rate, x_4 = pressure, x_5 = exhaust temperature, and x_6 = ambient temperature at time of test.

(a) Fit a multiple linear regression model using x_3 = fuel flow rate, x_4 = pressure, and x_5 = exhaust temperature as the regressors.
(b) Test for significance of regression using α = 0.01. Find the *P*-value for this test. What are your conclusions?
(c) Find the *t*-test statistic for each regressor. Using α = 0.01, explain carefully the conclusion you can draw from these statistics.
(d) Find R^2 and the adjusted statistic for this model.
(e) Construct a normal probability plot of the residuals and interpret this graph.
(f) Plot the residuals versus $\hat{y}$. Are there any indications of inequality of variance or nonlinearity?
(g) Plot the residuals versus x_3. Is there any indication of nonlinearity?
(h) Predict the thrust for an engine for which x_2 = 28900, x_4 = 170, and x_5 = 1589.

(a) $\hat{y} = 4203 - 0.231x_3 + 21.485x_4 + 1.6887x_5$

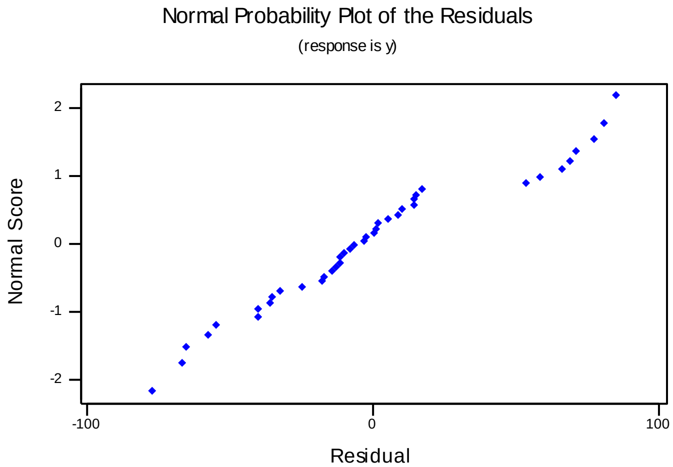
(b) $H_0: \beta_3 = \beta_4 = \beta_5 = 0$
$H_1: \beta_j \neq 0$ for at least one j
$\alpha = 0.01$ $f_0 = 1651.25$
$f_{.01,3,36} = 4.38$
Reject H_0 and conclude that regression is significant. P-value < 0.00001

(c) All at α = 0.01 $t_{.005,36} = 2.72$

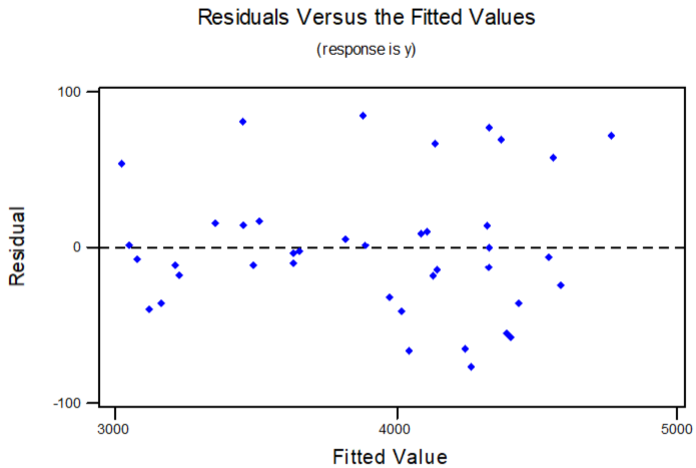
$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$	$H_0: \beta_5 = 0$
$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$	$H_1: \beta_5 \neq 0$
$t_0 = -2.06$	$t_0 = 22.91$	$t_0 = 3.00$
$\lvert t_0 \rvert \ngtr t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$	$\lvert t_0 \rvert > t_{\alpha/2,36}$
Fail to reject H_0	Reject H_0	Reject H_0

(d) $R^2 = 0.993$ Adj. $R^2 = 0.9925$

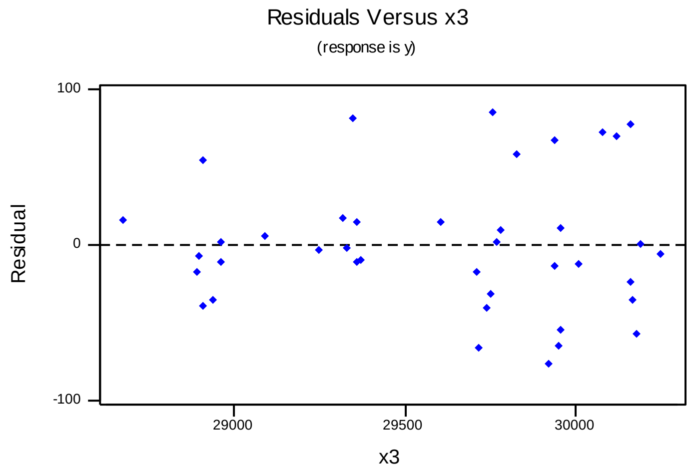
(e) Normality assumption appears reasonable. However there is a gap in the line.



(f) Plot is satisfactory.



(g) Slight indication that variance increases as x_3 increases.



(h) $\hat{y} = 4203 - 0.231(28900) + 21.485(170) + 1.6887(1589) = 3862.89$

12-109 Consider the engine thrust data in Exercise 12-108. Refit the model using $y^* = \ln y$ as the response variable and $x_3^* = \ln_3$ as the regressor (along with x_4 and x_5).

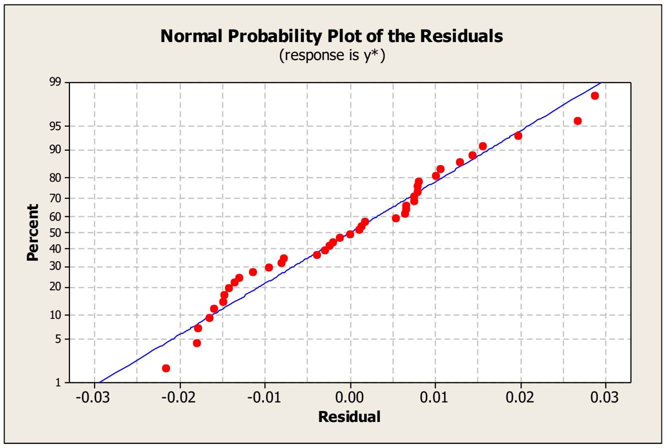
(a) Test for significance of regression using $\alpha = 0.01$. Find the *P*-value for this test and state your conclusions.
(b) Use the *t*-statistic to test $H_0\!:\beta_j = 0$ versus $H_1\!:\beta_j \neq 0$ for each variable in the model. If $\alpha = 0.01$, what conclusions can you draw?
(c) Plot the residuals versus $\hat{y}^*$ and versus x_3^*. Comment on these plots. How do they compare with their counterparts obtained in Exercise 12-108 parts (f) and (g)?

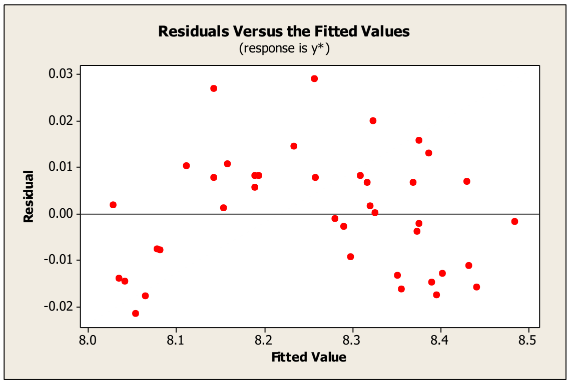
(a) $H_0\!:\beta_3^* = \beta_4 = \beta_5 = 0$
$H_1\!:\beta_j \neq 0$ for at least one j
$\alpha = 0.01$
$f_0 = 1321.39$
$f_{.01,3,36} = 4.38$
$f_0 >> f_{\alpha,3,36}$
Reject H_0 and conclude that regression is significant. P-value < 0.00001

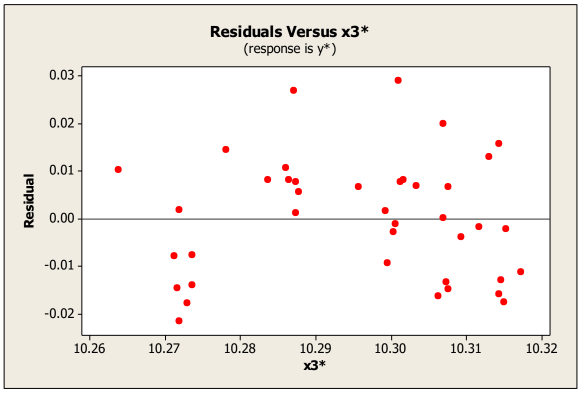
(b) $\alpha = 0.01$ $t_{.005,36} = 2.72$

$H_0\!:\beta_3^* = 0$	$H_0\!:\beta_4 = 0$	$H_0\!:\beta_5 = 0$
$H_1\!:\beta_3^* \neq 0$	$H_1\!:\beta_4 \neq 0$	$H_1\!:\beta_5 \neq 0$
$t_0 = -1.45$	$t_0 = 19.95$	$t_0 = 2.53$
$\lvert t_0\rvert \not> t_{\alpha/2,36}$	$\lvert t_0\rvert > t_{\alpha/2,36}$	$\lvert t_0\rvert \not> t_{\alpha/2,36}$
Fail to reject H_0	Reject H_0	Fail to reject H_0

(c) Curvature is evident in the residuals plots from this model.







12-110 Transient points of an electronic inverter are influenced by many factors. Table E12-15 gives data on the transient point (y, in volts) of PMOS-NMOS inverters and five candidate regressors: x_1 = width of the NMOS device, x_2 = length of the NMOS device, x_3 = width of the PMOS device, x_4 = length of the PMOS device, and x_5 = temperature (°C).

(a) Fit a multiple linear regression model that uses all regressors to these data. Test for significance of regression using $\alpha = 0.01$. Find the *P*-value for this test and use it to draw your conclusions.
(b) Test the contribution of each variable to the model using the *t*-test with $\alpha = 0.05$. What are your conclusions?
(c) Delete x_5 from the model. Test the new model for significance of regression. Also test the relative contribution of each regressor to the new model with the *t*-test. Using $\alpha = 0.05$, what are your conclusions?
(d) Notice that the MS_E for the model in part (c) is smaller than the MS_E for the full model in part (a). Explain why this has occurred.
(e) Calculate the studentized residuals. Do any of these seem unusually large?
(f) Suppose that you learn that the second observation was recorded incorrectly. Delete this observation and refit the model using x_1, x_2, x_3, and x_4 as the regressors. Notice that the R^2 for this model is considerably higher than the R^2 for either of the models fitted previously. Explain why the R^2 for this model has increased.
(g) Test the model from part (f) for significance of regression using $\alpha = 0.05$. Also investigate the contribution of each regressor to the model using the *t*-test with $\alpha = 0.05$. What conclusions can you draw?

TABLE • E12-14 Thrust of a Jet-Turbine Engine

Observation Number	y	x_1	x_2	x_3	x_4	x_5	x_6
1	4540	2140	20640	30250	205	1732	99
2	4315	2016	20280	30010	195	1697	100
3	4095	1905	19860	29780	184	1662	97
4	3650	1675	18980	29330	164	1598	97
5	3200	1474	18100	28960	144	1541	97
6	4833	2239	20740	30083	216	1709	87
7	4617	2120	20305	29831	206	1669	87
8	4340	1990	19961	29604	196	1640	87
9	3820	1702	18916	29088	171	1572	85
10	3368	1487	18012	28675	149	1522	85
11	4445	2107	20520	30120	195	1740	101
12	4188	1973	20130	29920	190	1711	100
13	3981	1864	19780	29720	180	1682	100
14	3622	1674	19020	29370	161	1630	100
15	3125	1440	18030	28940	139	1572	101
16	4560	2165	20680	30160	208	1704	98
17	4340	2048	20340	29960	199	1679	96
18	4115	1916	19860	29710	187	1642	94
19	3630	1658	18950	29250	164	1576	94
20	3210	1489	18700	28890	145	1528	94
21	4330	2062	20500	30190	193	1748	101
22	4119	1929	20050	29960	183	1713	100
23	3891	1815	19680	29770	173	1684	100
24	3467	1595	18890	29360	153	1624	99
25	3045	1400	17870	28960	134	1569	100
26	4411	2047	20540	30160	193	1746	99
27	4203	1935	20160	29940	184	1714	99
28	3968	1807	19750	29760	173	1679	99
29	3531	1591	18890	29350	153	1621	99
30	3074	1388	17870	28910	133	1561	99
31	4350	2071	20460	30180	198	1729	102
32	4128	1944	20010	29940	186	1692	101
33	3940	1831	19640	29750	178	1667	101
34	3480	1612	18710	29360	156	1609	101
35	3064	1410	17780	28900	136	1552	101
36	4402	2066	20520	30170	197	1758	100
37	4180	1954	20150	29950	188	1729	99
38	3973	1835	19750	29740	178	1690	99
39	3530	1616	18850	29320	156	1616	99
40	3080	1407	17910	28910	137	1569	100

TABLE • E12-15 Transient Point of an Electronic Inverter

Observation Number	x_1	x_2	x_3	x_4	x_5	y
1	3	3	3	3	0	0.787
2	8	30	8	8	0	0.293
3	3	6	6	6	0	1.710
4	4	4	4	12	0	0.203
5	8	7	6	5	0	0.806
6	10	20	5	5	0	4.713
7	8	6	3	3	25	0.607
8	6	24	4	4	25	9.107
9	4	10	12	4	25	9.210
10	16	12	8	4	25	1.365
11	3	10	8	8	25	4.554
12	8	3	3	3	25	0.293
13	3	6	3	3	50	2.252
14	3	8	8	3	50	9.167
15	4	8	4	8	50	0.694
16	5	2	2	2	50	0.379
17	2	2	2	3	50	0.485
18	10	15	3	3	50	3.345
19	15	6	2	3	50	0.208
20	15	6	2	3	75	0.201
21	10	4	3	3	75	0.329
22	3	8	2	2	75	4.966
23	6	6	6	4	75	1.362
24	2	3	8	6	75	1.515
25	3	3	8	8	75	0.751

(h) Plot the residuals from the model in part (f) versus $\hat{y}$ and versus each of the regressors x_1, x_2, x_3, and x_4. Comment on the plots.

(a) $\hat{y} = 2.86 - 0.291x_1 + 0.2206x_2 + 0.454x_3 - 0.594x_4 + 0.005x_5$

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: at least one $\beta_j \neq 0$

$\alpha = 0.01$

$f_0 = 4.81$

$f_{.01,5,19} = 4.17$

$f_0 > f_{\alpha,5,19}$

Reject H_0. *P*-value = 0.005

(b) $\alpha = 0.05$ $t_{.025,19} = 2.093$

$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$	$H_0: \beta_5 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$	$H_1: \beta_5 \neq 0$
$t_0 = -2.47$	$t_0 = 2.74$	$t_0 = 2.42$	$t_0 = -2.80$	$t_0 = 0.26$
$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert > t_{\alpha/2,19}$	$\lvert t_0 \rvert \ngtr t_{\alpha/2,19}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0	Fail to reject H_0

(c) $\hat{y} = 3.148 - 0.290x_1 + 0.19919x_2 + 0.455x_3 - 0.609x_4$

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_1: \beta_j \neq 0$ for at least one j

$\alpha = 0.05$

$f_0 = 6.28$

$f_{.05,4,20} = 2.87$

$f_0 > f_{\alpha,4,20}$

Reject H_0.

$\alpha = 0.05$ $t_{.025,20} = 2.086$

$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$
$t_0 = -2.53$	$t_0 = 2.89$	$t_0 = 2.49$	$t_0 = -3.05$
$\|t_0\| > t_{\alpha/2,20}$	$\|t_0\| > t_{\alpha/2,20}$	$\|t_0\| > t_{\alpha/2,20}$	$\|t_0\| > t_{\alpha/2,20}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0

(d) The addition of the 5th regressor results in a loss of one degree of freedom in the denominator and the reduction in SS_E is not enough to compensate for this loss.

(e) Observation 2 is unusually large. Studentized residuals follow:

-0.80199	-4.99898	-0.39958	2.22883	-0.52268	0.62842	-0.45288	2.21003	1.37196
-0.42875	0.75434	-0.32059	-0.36966	1.91673	0.08151	-0.69908	-0.79465	0.27842
0.59987	0.59609	-0.12247	0.71898	-0.73234	-0.82617	-0.45518		

(f) R^2 for model in part (a): 0.558. R^2 for model in part (c): 0.557. R^2 for model x_1, x_2, x_3, x_4 without obs. #2: 0.804. R^2 increased because observation 2 was not fit well by either of the previous models.

(g) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_1: \beta_j \neq 0$ $\alpha = 0.05$

$f_0 = 19.53$

$f_{.05,4,19} = 2.90$

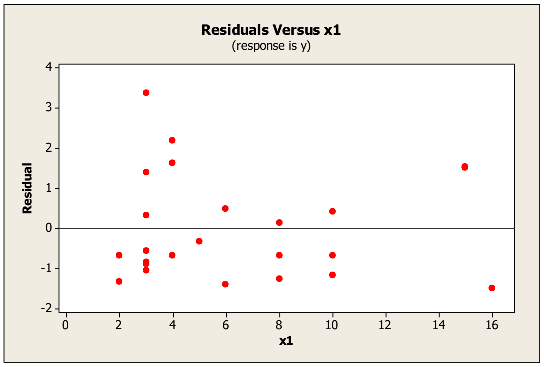
$f_0 > f_{0.05,4,19}$

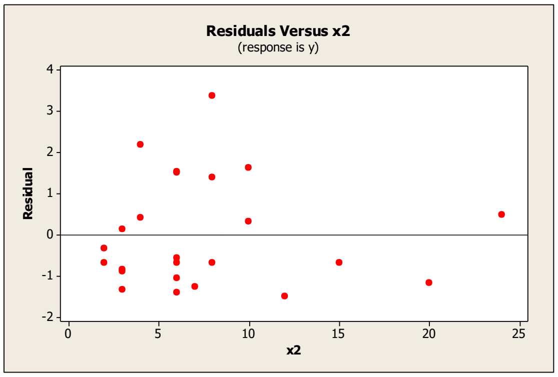
Reject H_0.

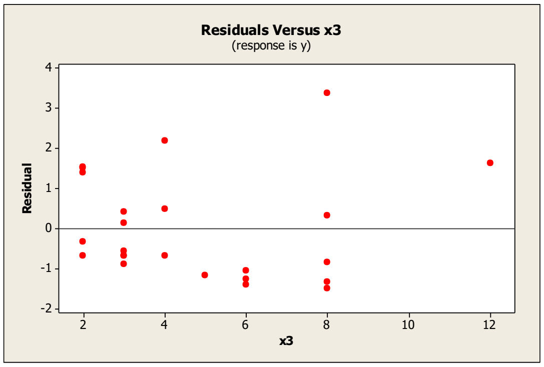
$\alpha = 0.05$ $t_{.025,19} = 2.093$

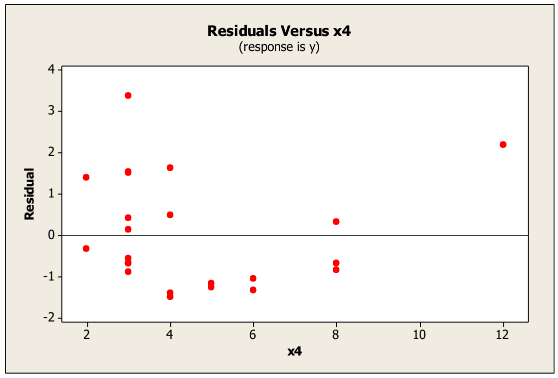
$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$	$H_0: \beta_3 = 0$	$H_0: \beta_4 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$
$t_0 = -3.96$	$t_0 = 6.43$	$t_0 = 3.64$	$t_0 = -3.39$
$\|t_0\| > t_{0.025,19}$	$\|t_0\| > t_{0.025,19}$	$\|t_0\| > t_{0.025,19}$	$\|t_0\| > t_{0.025,19}$
Reject H_0	Reject H_0	Reject H_0	Reject H_0

(h) There is some indication of curvature.









12-111 Consider the inverter data in Exercise 12-110. Delete observation 2 from the original data. Define new variables as follows: $y^* = \ln y$, $x_1^* = 1/\sqrt{x_1}$, $x_2^* = \sqrt{x_2}$, $x_3^* = 1/\sqrt{x_3}$, and $x_4^* = \sqrt{x_4}$.

(a) Fit a regression model using these transformed regressors (do not use x_5 or x_6).
(b) Test the model for significance of regression using α = 0.05. Use the *t*-test to investigate the contribution of each variable to the model (α = 0.05). What are your conclusions?
(c) Plot the residuals versus $\hat{y}^*$ and versus each of the transformed regressors. Comment on the plots.

Note that data in row 2 are deleted to follow the instructions in the exercise.

```
(a)
The regression equation is
  y* = - 0.908 + 5.48 x1* + 1.13 x2* - 3.92 x3*
             - 1.14 x4*

Predictor        Coef  SE Coef      T      P
Constant      -0.9082   0.6746  -1.35  0.194
  x1*          5.4823   0.4865  11.27  0.000
  x2*         1.12563  0.07714  14.59  0.000
  x3*         -3.9198   0.5619  -6.98  0.000
  x4*         -1.1429   0.1410  -8.11  0.000

S = 0.282333   R-Sq = 95.8%   R-Sq(adj) = 94.9%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       4  34.7600  8.6900  109.02  0.000
Residual Error  19   1.5145  0.0797
Total           23  36.2745
```

(b) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

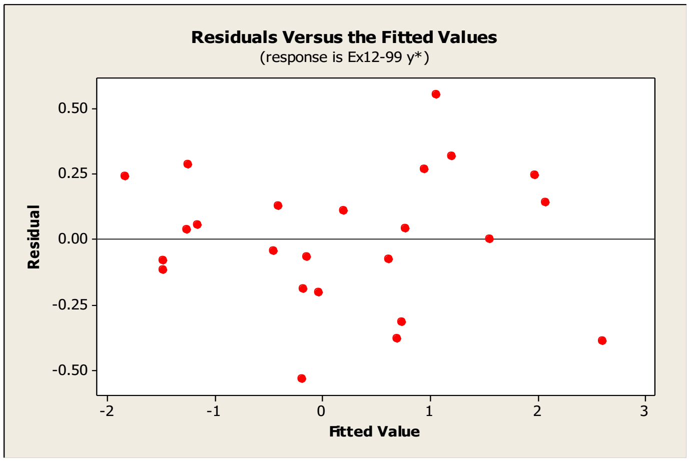
H_1 : at least one $\beta_j \neq 0$

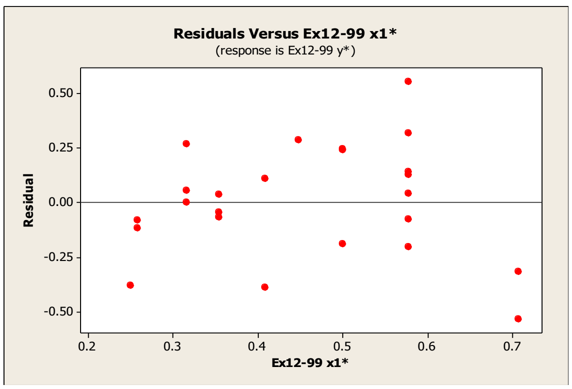
α = 0.05

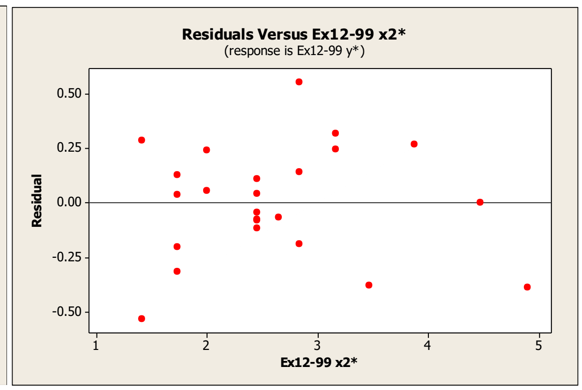
$f_0 = 109.02$, P-value ≈ 0,

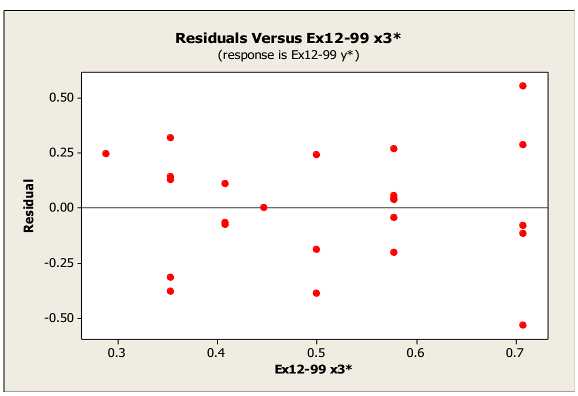
Reject H_0 at α = 0.05.
T tests appear in the previous computer output. Because all P-values ≈ 0, all tests reject H_0

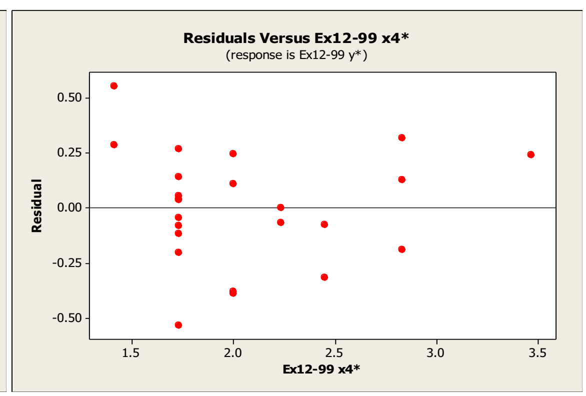
(c) The residual plots are more satisfactory than the plots in the previous exercise.











12-112 Following are data on y = green liquor (g/l) and x = paper machine speed (feet per minute) from a Kraft paper machine. (The data were read from a graph in an article in the *Tappi Journal*, March 1986.)

y	16.0	15.8	15.6	15.5	14.8
x	1700	1720	1730	1740	1750
y	14.0	13.5	13.0	12.0	11.0
x	1760	1770	1780	1790	1795

.

(a) Fit the model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \in$ using least squares.

(b) Test for significance of regression using $\alpha = 0.05$. What are your conclusions?

(c) Test the contribution of the quadratic term to the model, over the contribution of the linear term, using an *F*-statistic. If α = 0.05, what conclusion can you draw?
(d) Plot the residuals from the model in part (a) versus $\hat{y}$. Does the plot reveal any inadequacies?
(e) Construct a normal probability plot of the residuals. Comment on the normality assumption.

(a) $\hat{y} = -1709.405 + 2.023x - 0.0006x^2$

(b) $H_0 : \beta_1 = \beta_{11} = 0$

H_1 : at least one $\beta_j \neq 0$

$\alpha = 0.05$

$f_0 = 300.11$

$f_{.05,2,7} = 4.74$

$f_0 >> f_{0.05,2,7}$

Reject H_0.

(c) $H_0: \beta_{11} = 0$

$H_1: \beta_{11} \neq 0$

$\alpha = 0.05$

$$F_0 = \frac{SS_R(\beta_{11} \mid \beta_1)/r}{MS_E} = \frac{2.4276/1}{0.04413}$$

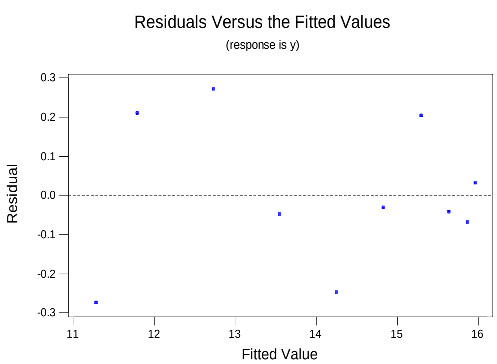
$f_0 = 55.01$

$f_{.05,1,7} = 5.59$

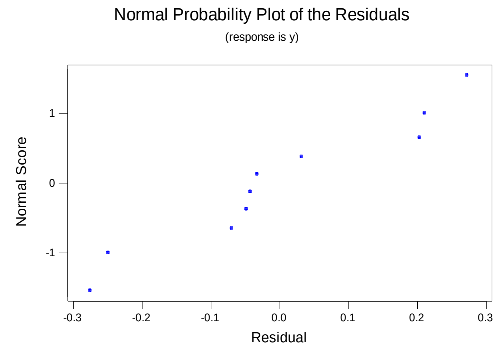
$f_0 >> f_{\alpha,1,7}$

Reject H_0.

(d) There is some indication of non-constant variance.



(e) Normality assumption is reasonable.



12-113 Consider the jet engine thrust data in Exercises 12-108 and 12-109. Define the response and regressors as in Exercise 12-109.

(a) Use all possible regressions to select the best regression equation, where the model with the minimum value of MS_E is to be selected as "best."
(b) Repeat part (a) using the C_p criterion to identify the best equation.
(c) Use stepwise regression to select a subset regression model.
(d) Compare the models obtained in parts (a), (b), and (c).
(e) Consider the three-variable regression model. Calculate the variance inflation factors for this model. Would you conclude that multicollinearity is a problem in this model?

(a) $\hat{y}^* = 21.068 - 1.404x_3^* + 0.0055x_4 + 0.000418x_5$

$MS_E = 0.013156$

$C_p = 4.0$

(b) Same as model in part (a)

(c) x_4, x_5 with $C_p = 4.1$ and $MS_E = 0.0134$

(d) The model in part (c) is simpler with values for MS_E and C_p similar to those in part (a) and (b). The part (c) model is preferable.

(e) $VIF(\hat{\beta}_3^*) = 52.4$

$VIF(\hat{\beta}_4) = 9.3$

$VIF(\hat{\beta}_5) = 29.1$

Yes, VIFs for x_3^* and x_5 exceed 10

12-114 Consider the electronic inverter data in Exercises 12-110 and 12-111. Define the response and regressors variables as in Exercise 12-111, and delete the second observation in the sample.

(a) Use all possible regressions to find the equation that minimizes C_p.
(b) Use all possible regressions to find the equation that minimizes MS_E.
(c) Use stepwise regression to select a subset regression model.
(d) Compare the models you have obtained.

(a) $\hat{y} = 4.87 + 6.12x_1^* - 6.53x_2^* - 3.56x_3^* - 1.44x_4^*$

$MS_E(p) = 0.41642$

Min $C_p = 5.0$

(b) Same as part (a)
(c) Same as part (a)
(d) All models are the same.

12-115 A multiple regression model was used to relate y = viscosity of a chemical product to x_1 = temperature and x_2 = reaction time. The data set consisted of $n = 15$ observations.
(a) The estimated regression coefficients were $\hat{\beta}_0 = 300.00$, $\hat{\beta}_1 = 0.85$, and $\hat{\beta}_2 = 10.40$. Calculate an estimate of mean viscosity when $x_1 = 100$°F and $x_2 = 2$ hours.
(b) The sums of squares were $SS_T = 1230.50$ and $SS_E = 120.30$. Test for significance of regression using $\alpha = 0.05$. What

conclusion can you draw?
(c) What proportion of total variability in viscosity is accounted for by the variables in this model?
(d) Suppose that another regressor, x_3 = stirring rate, is added to the model. The new value of the error sum of squares is SS_E = 117.20. Has adding the new variable resulted in a smaller value of MS_E? Discuss the significance of this result.
(e) Calculate an F-statistic to assess the contribution of x_3 to the model. Using α = 0.05, what conclusions do you reach?

(a) $\hat{y} = 300.0 + 0.85x_1 + 10.4x_2$
$\hat{y} = 300 + 0.85(100) + 10.4(2) = 405.8$

(b) $S_{yy} = 1230.5$ $SS_E = 120.3$

$$SS_R = S_{yy} - SS_E = 1230.5 - 120.3 = 1110.2$$

$$MS_R = \frac{SS_R}{k} = \frac{1110.2}{2} = 555.1$$

$$MS_E = \frac{SS_E}{n-p} = \frac{120.3}{15-3} = 10.025$$

$$f_0 = \frac{MS_R}{MS_E} = \frac{555.1}{10.025} = 55.37$$

$H_0: \beta_1 = \beta_2 = 0$
H_1 : at least one $\beta_j \neq 0$
α = 0.05
$f_0 = 55.37$
$f_{.05,2,12} = 3.89$
$f_0 > f_{0.05,2,12}$
Reject H_0 and conclude that the regression model is significant at α = 0.05

(c) $R^2 = \frac{SS_R}{S_{yy}} = \frac{1110.2}{1230.5} = 0.9022$ or 90.22%

(d) k = 3 p = 4 $SS_E' = 117.20$

$$MS_E' = \frac{SS_E'}{n-p} = \frac{117.2}{11} = 10.65$$

No, MS_E increased with the addition of x_3 because the reduction in SS_E was not enough to compensate for the loss in one degree of freedom in the error sum of squares. This is why MS_E can be used as a model selection criterion.

(e) $SS_R = S_{yy} - SS_E = 1230.5 - 117.20 = 1113.30$

$$SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0) = SS_R(\beta_3\beta_2\beta_1 \mid \beta_0) - SS_R(\beta_2, \beta_1 \mid \beta_0)$$
$$= 1113.30 - 1110.20$$
$$= 3.1$$

$H_0: \beta_3 = 0$
$H_1: \beta_3 \neq 0$
α = 0.05

$$f_0 = \frac{SS_R(\beta_3 \mid \beta_2, \beta_1, \beta_0)/r}{SS_E'/n-p} = \frac{3.1/1}{117.2/11} = 0.291$$

$f_{.05,1,11} = 4.84$
$f_0 \not> f_{0.05,1,11}$
Fail to reject H_0.

12-116 Tables E12-16 and E12-17 present statistics for the Major League Baseball 2005 season (*The Sports Network*).

(a) Consider the batting data. Use model-building methods to predict *wins* from the other variables. Check that the assumptions for your model are valid.
(b) Repeat part (a) for the pitching data.
(c) Use both the batting and pitching data to build a model to predict *wins.* What variables are most important? Check that the assumptions for your model are valid.

TABLE • E12-16 Major League Baseball 2005 Season

	American League Batting													
Team	**W**	**AVG**	**R**	**H**	**2B**	**3B**	**HR**	**RBI**	**BB**	**SO**	**SB**	**GIDP**	**LOB**	**OBP**
Chicago	99	0.262	741	1450	253	23	200	713	435	1002	137	122	1032	0.322
Boston	95	0.281	910	1579	339	21	199	863	653	1044	45	135	1249	0.357
LA Angels	95	0.27	761	1520	278	30	147	726	447	848	161	125	1086	0.325
New York	95	0.276	886	1552	259	16	229	847	637	989	84	125	1264	0.355
Cleveland	93	0.271	790	1522	337	30	207	760	503	1093	62	128	1148	0.334
Oakland	88	0.262	772	1476	310	20	155	739	537	819	31	148	1170	0.33
Minnesota	83	0.259	688	1441	269	32	134	644	485	978	102	155	1109	0.323
Toronto	80	0.265	775	1480	307	39	136	735	486	955	72	126	1118	0.331
Texas	79	0.267	865	1528	311	29	260	834	495	1112	67	123	1104	0.329
Baltimore	74	0.269	729	1492	296	27	189	700	447	902	83	145	1103	0.327
Detroit	71	0.272	723	1521	283	45	168	678	384	1038	66	137	1077	0.321
Seattle	69	0.256	699	1408	289	34	130	657	466	986	102	115	1076	0.317
Tampa Bay	67	0.274	750	1519	289	40	157	717	412	990	151	133	1065	0.329
Kansas City	56	0.263	701	1445	289	34	126	653	424	1008	53	139	1062	0.32

	National League Batting													
Team	**W**	**AVG**	**R**	**H**	**2B**	**3B**	**HR**	**RBI**	**BB**	**SO**	**SB**	**GIDP**	**LOB**	**OBP**
St. Louis	100	0.27	805	1494	287	26	170	757	534	947	83	127	1152	0.339
Atlanta	90	0.265	769	1453	308	37	184	733	534	1084	92	146	1114	0.333
Houston	89	0.256	693	1400	281	32	161	654	481	1037	115	116	1136	0.322
Philadelphia	88	0.27	807	1494	282	35	167	760	639	1083	116	107	1251	0.348
Florida	83	0.272	717	1499	306	32	128	678	512	918	96	144	1181	0.339
New York	83	0.258	722	1421	279	32	175	683	486	1075	153	103	1122	0.322
San Diego	82	0.257	684	1416	269	39	130	655	600	977	99	122	1220	0.333
Milwaukee	81	0.259	726	1413	327	19	175	689	531	1162	79	137	1120	0.331
Washington	81	0.252	639	1367	311	32	117	615	491	1090	45	130	1137	0.322
Chicago	79	0.27	703	1506	323	23	194	674	419	920	65	131	1133	0.324
Arizona	77	0.256	696	1419	291	27	191	670	606	1094	67	132	1247	0.332
San Francisco	75	0.261	649	1427	299	26	128	617	431	901	71	147	1093	0.319
Cincinnati	73	0.261	820	1453	335	15	222	784	611	1303	72	116	1176	0.339
Los Angeles	71	0.253	685	1374	284	21	149	653	541	1094	58	139	1135	0.326
Colorado	67	0.267	740	1477	280	34	150	704	509	1103	65	125	1197	0.333
Pittsburgh	67	0.259	680	1445	292	38	139	656	471	1092	73	130	1193	0.322

Batting

W	Wins
AVG	Batting average
R	Runs
H	Hits
2B	Doubles
3B	Triples
HR	Home runs
RBI	Runs batted in
BB	Walks
SO	Strikeouts
SB	Stolen bases
GIDP	Grounded into double play
LOB	Left on base
OBP	On-base percentage

Pitching

ERA	Earned run average
SV	Saves
H	Hits
R	Runs
ER	Earned runs
HR	Home runs
BB	Walks
SO	Strikeouts
AVG	Opponent batting average

TABLE • E12-17 Major League Baseball 2005

	American League Pitching									
Team	**W**	**ERA**	**SV**	**H**	**R**	**ER**	**HR**	**BB**	**SO**	**AVG**
Chicago	99	3.61	54	1392	645	592	167	459	1040	0.249
Boston	95	4.74	38	1550	805	752	164	440	959	0.276
LA Angels	95	3.68	54	1419	643	598	158	443	1126	0.254
New York	95	4.52	46	1495	789	718	164	463	985	0.269
Cleveland	93	3.61	51	1363	642	582	157	413	1050	0.247
Oakland	88	3.69	38	1315	658	594	154	504	1075	0.241
Minnesota	83	3.71	44	1458	662	604	169	348	965	0.261
Toronto	80	4.06	35	1475	705	653	185	444	958	0.264
Texas	79	4.96	46	1589	858	794	159	522	932	0.279
Baltimore	74	4.56	38	1458	800	724	180	580	1052	0.263
Detroit	71	4.51	37	1504	787	719	193	461	907	0.272
Seattle	69	4.49	39	1483	751	712	179	496	892	0.268
Tampa Bay	67	5.39	43	1570	936	851	194	615	949	0.28
Kansas City	56	5.49	25	1640	935	862	178	580	924	0.291
	National League Pitching									
Team	**W**	**ERA**	**SV**	**H**	**R**	**ER**	**HR**	**BB**	**SO**	**AVG**
St. Louis	100	3.49	48	1399	634	560	153	443	974	0.257
Atlanta	90	3.98	38	1487	674	639	145	520	929	0.268
Houston	89	3.51	45	1336	609	563	155	440	1164	0.246
Philadelphia	88	4.21	40	1379	726	672	189	487	1159	0.253
Florida	83	4.16	42	1459	732	666	116	563	1125	0.266
New York	83	3.76	38	1390	648	599	135	491	1012	0.255
San Diego	82	4.13	45	1452	726	668	146	503	1133	0.259
Milwaukee	81	3.97	46	1382	697	635	169	569	1173	0.251
Washington	81	3.87	51	1456	673	627	140	539	997	0.262
Chicago	79	4.19	39	1357	714	671	186	576	1256	0.25
Arizona	77	4.84	45	1580	856	783	193	537	1038	0.278
San Francisco	75	4.33	46	1456	745	695	151	592	972	0.263
Cincinnati	73	5.15	31	1657	889	820	219	492	955	0.29
Los Angeles	71	4.38	40	1434	755	695	182	471	1004	0.263
Colorado	67	5.13	37	1600	862	808	175	604	981	0.287
Pittsburgh	67	4.42	35	1456	769	706	162	612	958	0.267

Batting

W	Wins
AVG	Batting average
R	Runs
H	Hits
2B	Doubles
3B	Triples
HR	Home runs
RBI	Runs batted in
BB	Walks
SO	Strikeouts
SB	Stolen bases
GID	Grounded into double play
LOB	Left on base
OBP	On-base percentage

Pitching

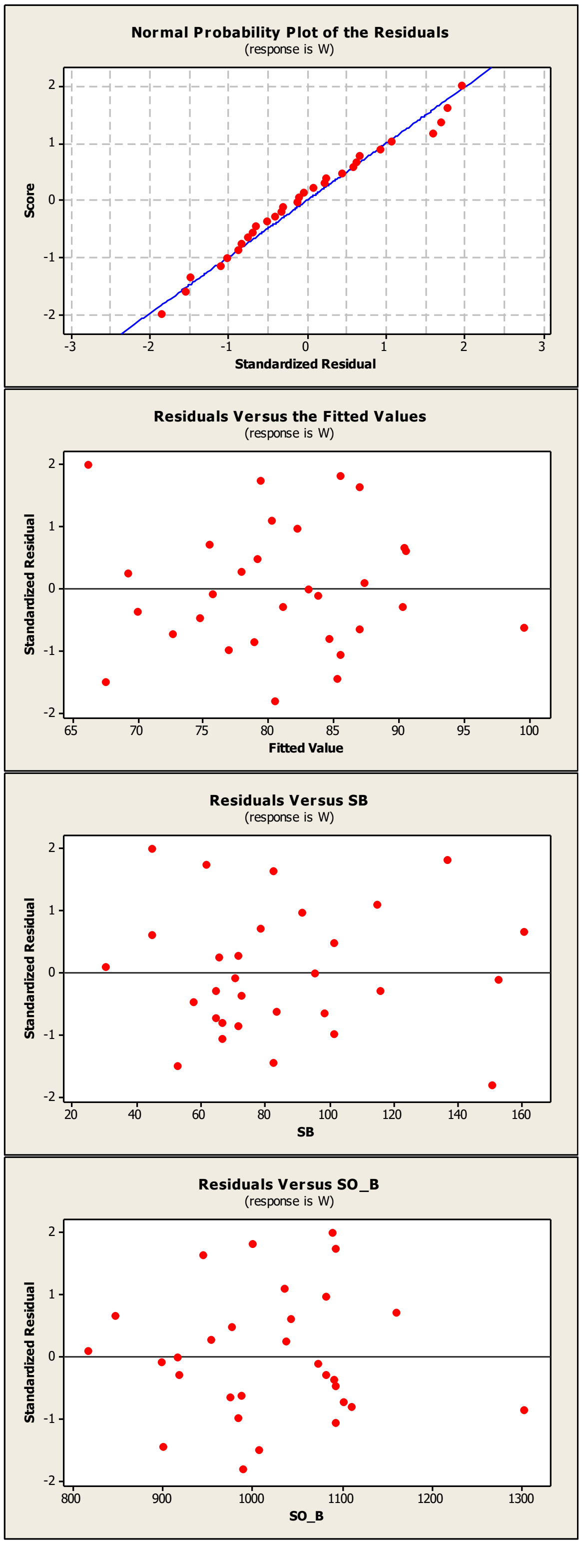
ERA	Earned run average
SV	Saves
H	Hits
R	Runs
ER	Earned runs
HR	Home runs
BB	Walks
SO	Strikeouts
AVG	Opponent batting average

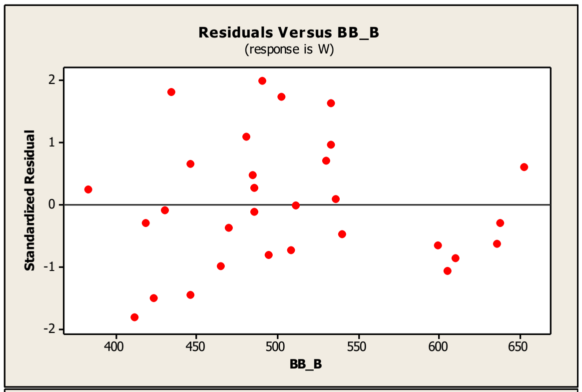
(a) The model with the minimum C_p (-1.3) value is:

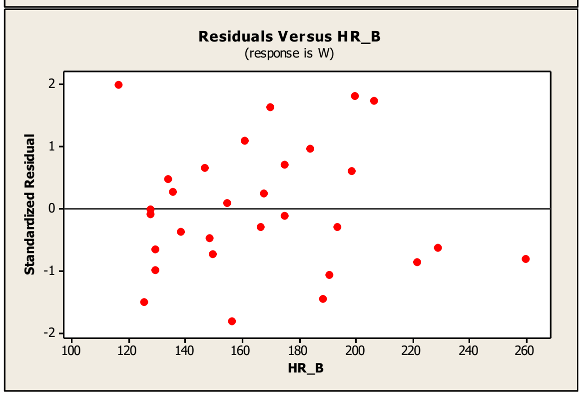
```
W = 69.9 + 0.120 HR_B + 0.0737 BB_B - 0.0532 SO_B + 0.0942 SB
```

where X_1 = AVG, X_2 = R, X_3 = H, X_4 = 2B, X_5 = 3B, X_6 = HR, X_7 = RBI, X_8 = BB, X_9 = SO, X_{10} = SB, X_{11} = GIDP, X_{12} = LOB and X_{13} = OBP

The model assumptions are not violated based on the following graphs.

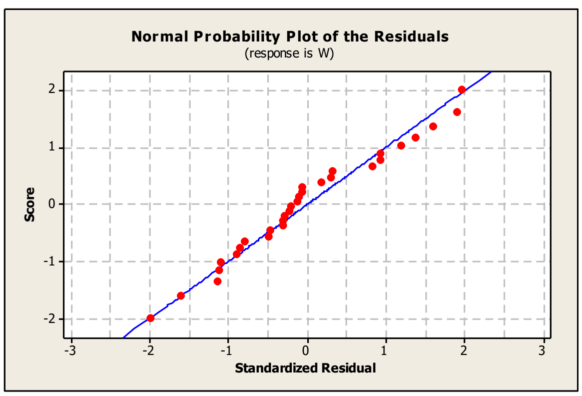

Normal Probability Plot of the Residuals
(response is W)
Score
Standardized Residual
Residuals Versus the Fitted Values
(response is W)
Standardized Residual
Fitted Value
Residuals Versus SB
(response is W)
Standardized Residual
SB
Residuals Versus SO_B
(response is W)
Standardized Residual
SO_B

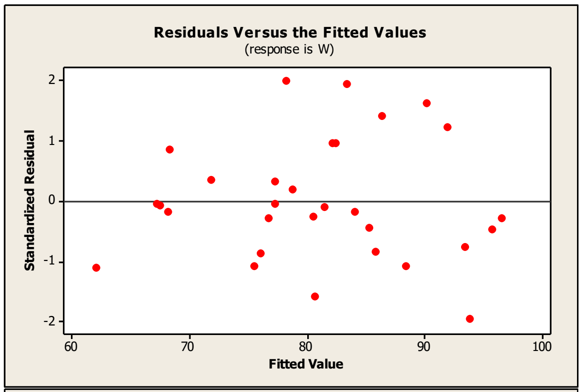


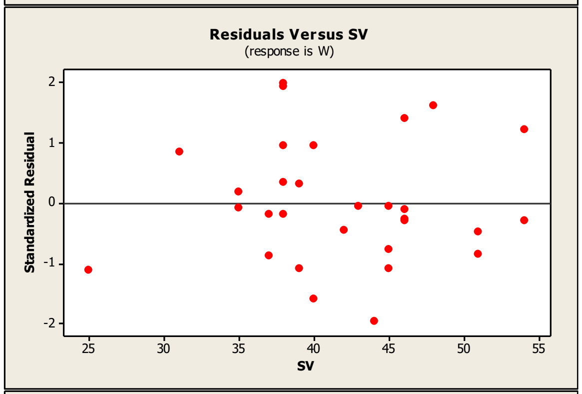


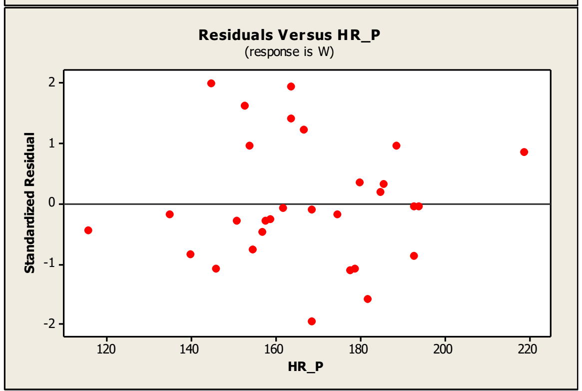
(b)
Minimum C_p (1.1) model:
`W = 96.5 + 0.527 SV - 0.125 HR_P - 0.0847 BB_P + 0.0257 SO_P`

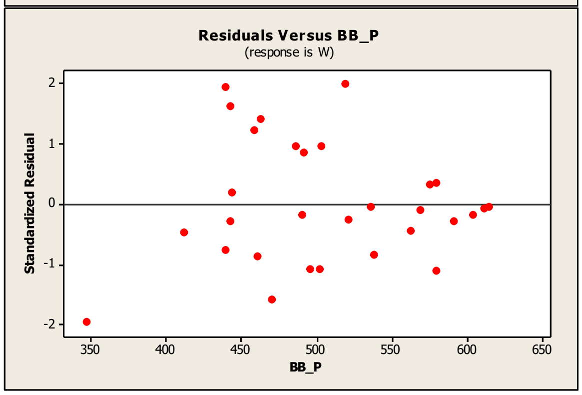
Based on the graphs below, the model assumptions are not violated

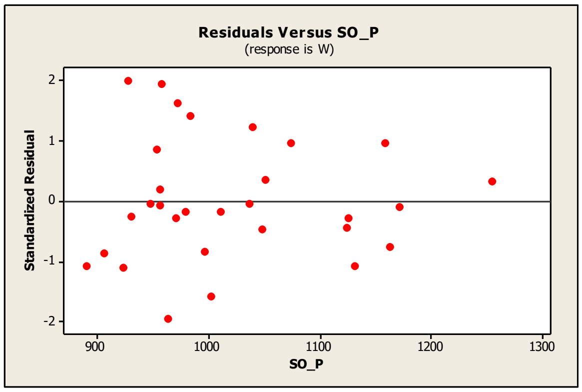



Residuals Versus the Fitted Values
(response is W)
Standardized Residual
Fitted Value


Residuals Versus SV
(response is W)
Standardized Residual
SV


Residuals Versus HR_P
(response is W)
Standardized Residual
HR_P

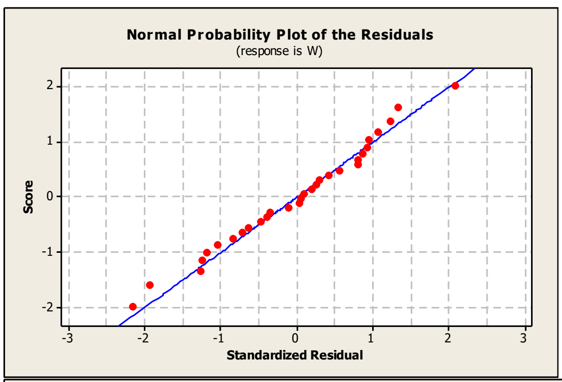

Residuals Versus BB_P
(response is W)
Standardized Residual
BB_P

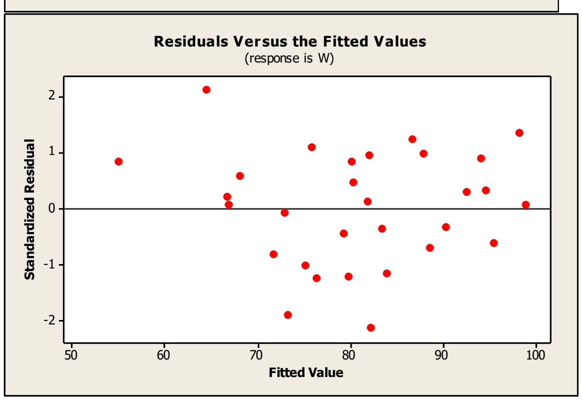


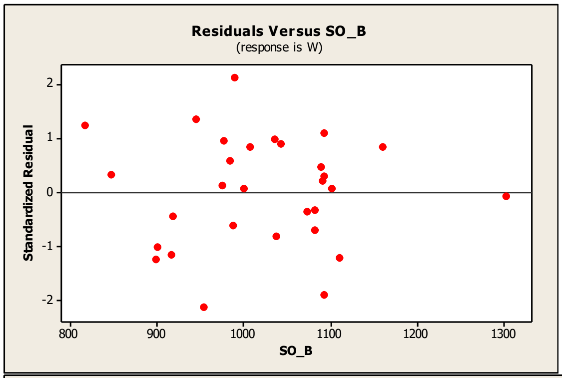
(c) Minimum C_p (10.7) model:

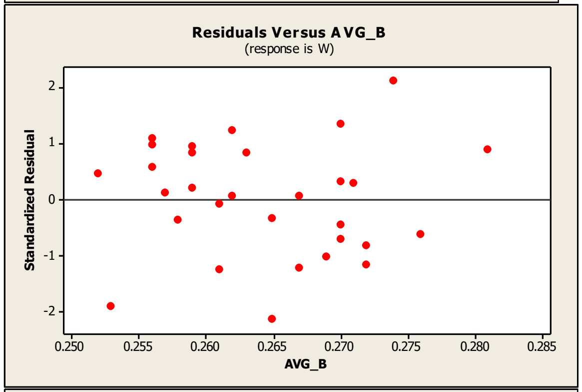
$$\hat{y} = -2.49 + 3277x_{avg\_b} - 0.45303x_{h\_b} - 0.04041x_{2b} - 0.13662x_{3b} + 0.19914x_{rbi}$$
$$- 0.010207x_{so\_b} + 0.07897x_{lob} - 870.2x_{obp} - 134.79x_{era} + 0.81681x_{er} - 0.06698x_{hr\_p}$$
$$- 0.032314x_{bb\_p} + 0.008755x_{so\_p}$$

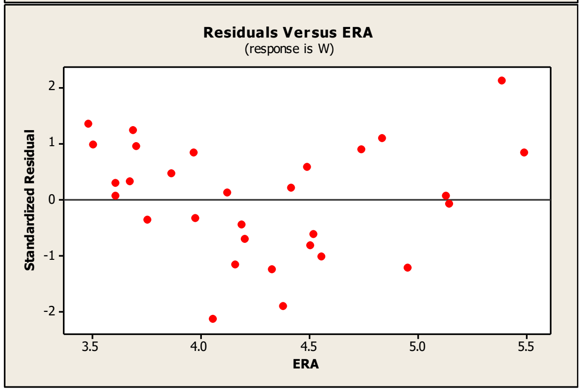
Every variable in the above model is significant at α = 0.10. If α is decreased to 0.05, SO_P is no longer significant. The residual plots do not show any violations of the model assumptions (only a few plots of residuals vs. the regressors are shown).











12-117 An article in the *Journal of the American Ceramics Society* (1992, Vol. 75, pp. 112–116) described a process for immobilizing chemical or nuclear wastes in soil by dissolving the contaminated soil into a glass block. The authors mix CaO and Na_2O with soil and model viscosity and electrical conductivity. The electrical conductivity model involves six regressors, and the sample consists of $n = 14$ observations.

(a) For the six-regressor model, suppose that $SS_T = 0.50$ and $R^2 = 0.94$. Find SS_E and SS_R, and use this information to test for significance of regression with $\alpha = 0.05$. What are your conclusions?
(b) Suppose that one of the original regressors is deleted from the model, resulting in $R^2 = 0.92$. What can you conclude about the contribution of the variable that was removed? Answer this question by calculating an F-statistic.
(c) Does deletion of the regressor variable in part (b) result in a smaller value of MS_E for the five-variable model, in comparison to the original six-variable model? Comment on the significance of your answer.

(a) $R^2 = \dfrac{SS_R}{S_{yy}}$

$$SS_R = R^2(S_{yy}) = 0.94(0.50) = 0.47$$
$$SS_E = S_{yy} - SS_R = 0.5 - 0.47 = 0.03$$
$$H_0: \beta_1 = \beta_2 = ... = \beta_6 = 0$$
$$H_1: \text{at least one } \beta_j \neq 0$$
$\alpha = 0.05$
$$f_0 = \frac{SS_R / k}{SS_E / n-p} = \frac{0.47/6}{0.03/7} = 18.28$$
$$f_{.05,6,7} = 3.87$$
$$f_0 > f_{0.05,6,7}$$
Reject H_0.

(b) $k = 5$ $n = 14$ $p = 6$ $R^2 = 0.92$
$$SS_R' = R^2(S_{yy}) = 0.92(0.50) = 0.46$$
$$SS_E' = S_{yy} - SS_R' = 0.5 - 0.46 = 0.04$$
$$SS_R(\beta\beta_{j_0}, \beta_{i,i=1,2,\ldots,6} \mid \beta_0) = SS_R(full) - SS_R(reduced)$$
$$= 0.47 - 0.46$$
$$= 0.01$$
$$f_0 = \frac{SS_R(\beta_j, \beta_{i,i=1,2,\ldots,6} \mid \beta_0)/r}{SS_E'/(n-p)} = \frac{0.01/1}{0.04/8} = 2$$
$$f_{.05,1,8} = 5.32$$
$$f_0 \not> f_{0.05,1,8}$$
Fail to reject H_0. There is not sufficient evidence that the removed variable is significant at $\alpha = 0.05$.

(c) $MS_E(reduced) = \dfrac{SS_E}{n-p} = \dfrac{0.04}{8} = 0.005$

$$MS_E(full) = \frac{0.03}{7} = 0.004$$
No, the MS_E is larger for the reduced model, although not by much. Generally, if adding a variable to a model reduces the MS_E it is an indication that the variable may be useful in explaining the response variable. Here the decrease in MS_E is not large because the added variable had no real explanatory power.

12-118 Exercise 12-9 introduced the hospital patient satisfaction survey data. One of the variables in that data set is a categorical variable indicating whether the patient is a medical patient or a surgical patient. Fit a model including this indicatorvariable to the data using all three of the other regressors. Is there any evidence that the service the patient is on (medical versus surgical) has an impact on the reported satisfaction?

The computer output is shown below. The P-value of the *Surg-Med* indicator variable (third variable) is greater than the $\alpha = 0.05$, so we fail to reject the H_0 and conclude that *Surg-Med* indicator variable does not contribute significantly to the model. Thus, the surgical and medical service does not impact the reported satisfaction.

Regression Analysis: Satisfaction versus Age, Severity, ...

```
The regression equation is
Satisfaction = 144 - 1.12 Age - 0.586 Severity + 0.41 Surg-Med + 1.31 Anxiety

Predictor      Coef  SE Coef      T      P
```

```
Constant   143.867    6.044  23.80  0.000
Age        -1.1172   0.1383  -8.08  0.000
Severity   -0.5862   0.1356  -4.32  0.000
Surg-Med     0.415    3.008   0.14  0.892
Anxiety      1.306    1.084   1.21  0.242

S = 7.20745   R-Sq = 90.4%   R-Sq(adj) = 88.4%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       4   9739.3  2434.8  46.87  0.000
Residual Error  20   1038.9    51.9
Total           24  10778.2
```

12-119 Consider the following inverse model matrix.

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 0.125 & 0 & 0 & 0 \\ 0 & 0.125 & 0 & 0 \\ 0 & 0 & 0.125 & 0 \\ 0 & 0 & 0 & 0.125 \end{bmatrix}$$

(a) How many regressors are in this model?
(b) What was the sample size?
(c) Notice the special diagonal structure of the matrix. What does that tell you about the columns in the original **X** matrix?

(a) Because the matrix is 4 by 4 there are 3 regressors in the model (plus the intercept).

(b) Because $(X'X)^{-1}$ is diagonal each element can be inverted to obtain $(X'X)$ and from the normal equations the (1, 1) element (the top-left element) of $(X'X)$ = n. Therefore, n = 1/0.125 = 8.

(c) The original columns are orthogonal to each other.

Mind-Expanding Exercises

12-120 Consider a multiple regression model with *k* regressors. Show that the test statistic for significance of regression can be written as

$$F_0 = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Suppose that $n = 20$, $k = 4$, and $R^2 = 0.90$. If $\alpha = 0.05$, what conclusion would you draw about the relationship between and the four regressors?

Because $R^2 = \frac{SS_R}{S_{yy}}$ and $1 - R^2 = \frac{SS_E}{S_{yy}}$,

$F_0 = \frac{SS_R/k}{SS_E/(n-k-1)}$ and this is the usual F-test for significance of regression. Then,

$F_0 = \frac{0.90/4}{(1-0.9)/(20-4-1)} = 33.75$ and the critical value is $f_{0.05,4,15} = 3.06$.

Because 33.75 > 3.06, regression is significant.

12-121 A regression model is used to relate a response to k = 4 regressors with n = 20. What is the smallest value of R^2 that will result in a significant regression if α = 0.05? Use the results of the previous exercise. Are you surprised by how small the value of R^2 is?

Using n = 20, k = 4, $f_{.05,4,15} = 3.06$. Reject H_0 if

$$\frac{R^2/4}{(1-R^2)/15} \geq 3.06$$

$$\frac{R^2}{(1-R^2)} \geq 0.816$$

Then, $R^2 \geq 0.449$ results in a significant regression.

12-122 Show that can express the residuals from a multiple regression model as $e = (\mathrm{I}-\mathrm{H})\mathrm{y}$ where $\mathrm{H} = \mathrm{X}(\mathrm{XX})^{-1}\mathrm{X}'$.

Because $\hat{\beta} = (X'X)^{-1}X'Y$, $e = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = (I-H)Y$

12-123 Show that the variance of the ith residual e_i in a multiple regression model is $\sigma^2(1-h_{ii})$ and that the covariance between e_i and e_j is $-\sigma^2 h_{ij}$ where the h's are the elements of $\mathrm{H} = \mathrm{X}(\mathrm{XX})^{-1}\mathrm{X}'$.

From the previous exercise, e_i is ith element of $(I-H)Y$. That is,

$$e_i = -h_{i,1}Y_1 - h_{i,2}Y_2 - ... - h_{i,i-1}Y_{i-1} + (1-h_{i,i})Y_i - h_{i,i+1}Y_{i+1} - ... - h_{i,n}Y_n$$

and

$$V(e_i) = (h_{i,1}^2 + h_{i,2}^2 + ... + h_{i,i-1}^2 + (1-h_{i,i}^2) + h_{i,i+1}^2 + ... + h_{i,n}^2)\sigma^2$$

The expression in parentheses is recognized to be the ith diagonal element of $(I-H)(I-H') = I-H$ by matrix multiplication. Consequently, $V(e_i) = (1-h_{i,i})\sigma^2$. Assume that $i < j$. Now,

$$e_i = -h_{i,1}Y_1 - h_{i,2}Y_2 - ... - h_{i,i-1}Y_{i-1} + (1-h_{i,i})Y_i - h_{i,i+1}Y_{i+1} - ... - h_{i,n}Y_n$$
$$e_j = -h_{j,1}Y_1 - h_{j,2}Y_2 - ... - h_{j,j-1}Y_{j-1} + (1-h_{j,j})Y_j - h_{j,j+1}Y_{j+1} - ... - h_{j,n}Y_n$$

Because the y_i's are independent,

$$Cov(e_i, e_j) = (h_{i,1}h_{j,1} + h_{i,2}h_{j,2} + ... + h_{i,i-1}h_{j,i-1} + (1-h_{i,i})h_{j,i}$$
$$+ h_{i,i+1}h_{j,i+1} + ... + h_{i,j}(1-h_{j,j}) + h_{i,j+1}h_{j,j+1} + ... + h_{i,n}h_{j,n})\sigma^2$$

The expression in parentheses is recognized as the ijth element of $(I-H)(I-H') = I-H$.
Therefore, $Cov(e_i, e_j) = -h_{ij}\sigma^2$.

12-124 Consider the multiple linear regression model $\mathrm{y} = \mathrm{X}\beta + \in$. If $\hat{\beta}$ denotes the least squares estimator of β , show that $\beta = \beta + \mathrm{R}\in$, where $\mathrm{R} = \mathrm{X}(\mathrm{X'X})^{-1}\mathrm{X}'$.

$$\hat{\beta} = (X'X)^{-1}X'Y = (\mathrm{X'X})^{-1}X'(X\beta + \varepsilon) = \beta + (\mathrm{X'X})^{-1}X'\varepsilon = \beta + R\varepsilon$$

12-125 **Constrained Least Squares.** Suppose we wish to find the least squares estimator of β in the model $\mathrm{y} = \mathrm{X}\beta + \in$. subject to a set of equality constraints, say, $\mathrm{T}\beta = \mathrm{c}$.

(a) Show that the estimator is

$$\hat{\beta}_c = \hat{\beta} + (\mathrm{X'X})^{-1} \times \mathrm{T}'\left[\mathrm{T}\left((\mathrm{X}'\mathrm{X})^{-1}\right)\mathrm{T}'\right]^{-1}(\mathrm{c} - \mathrm{T}\hat{\beta}) \text{ where } \hat{\beta} = (\mathrm{X'X})^{-1}\mathrm{X}'\mathrm{y}.$$

(b) Discuss situations where this model might be appropriate.

(a) Min L = $(\mathbf{y} - \mathbf{X\beta})'(\mathbf{y} - \mathbf{X\beta})$ subject to $\mathbf{T\beta = c}$

This is equivalent to Min Z = $(\mathbf{y} - \mathbf{X\beta})'(\mathbf{y} - \mathbf{X\beta}) + 2\boldsymbol{\gamma}'(\mathbf{T\beta} - \mathbf{c})$
where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots \gamma_p)'$ is a vector of Lagrange multipliers.

$$\frac{\partial Z}{\partial \beta} = -2X'y + 2(X'X)\beta + 2T'\gamma$$

$$\frac{\partial Z}{\partial \gamma} = 2(T\beta - c)$$

Set $\frac{\partial Z}{\partial \beta} = 0$ and $\frac{\partial Z}{\partial \gamma} = 0$

Then
$(\mathbf{X'X})\boldsymbol{\beta}_c + \mathbf{T}'\boldsymbol{\gamma} = \mathbf{X'y}$
$\mathbf{T}\boldsymbol{\beta}_c = \mathbf{c}$
where $\boldsymbol{\beta}_c$ is the constrained estimator.

From the first of these equations
$\boldsymbol{\beta}_c = (\mathbf{X'X})^{-1}(\mathbf{X'y} - \mathbf{T}'\boldsymbol{\gamma}) = \boldsymbol{\beta} - (\mathbf{X'X})^{-1}\mathbf{T}'\boldsymbol{\gamma}$

From the second
$\mathbf{T}\boldsymbol{\beta} - \mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T}'\boldsymbol{\gamma} = \mathbf{c}$ and $\boldsymbol{\gamma} = [\,\mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T}'\,]^{-1}(\mathbf{T}\boldsymbol{\beta} - \mathbf{c})$

Then
$\boldsymbol{\beta}_c = \boldsymbol{\beta} - (\mathbf{X'X})^{-1}\mathbf{T}'[\,\mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T}'\,]^{-1}(\mathbf{T}\boldsymbol{\beta} - \mathbf{c}) = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{T}'[\,\mathbf{T}(\mathbf{X'X})^{-1}\mathbf{T}'\,]^{-1}(\mathbf{c} - \mathbf{T}\boldsymbol{\beta})$

(b) This solution would be appropriate in situations where you know that there are linear relationships between the coefficients.

12-126 **Piecewise Linear Regression.** Suppose that y is piecewise linearly related to x. That is, different linear relationships are appropriate over the intervals $-\infty < x \le x^*$ and $x^* < x < \infty$.

(a) Show how indicator variables can be used to fit such a piecewise linear regression model, assuming that the point x^* is known.
(b) Suppose that at the point x^* a discontinuity occurs in the regression function. Show how indicator variables can be used to incorporate the discontinuity into the model.
(c) Suppose that the point x^* is not known with certainty and must be estimated. Suggest an approach that could be used to fit the piecewise linear regression model.

(a) For the piecewise linear function to be continuous at x = x*, the point-slope formula for a line can be used to show that

$$y = \begin{cases} \beta_0 + \beta_1(x - x^*) & x \le x^* \\ \beta_0 + \beta_2(x - x^*) & x > x^* \end{cases}$$

where $\beta_0, \beta_1, \beta_2$ are arbitrary constants.

Let $$z = \begin{cases} 0, & x \le x^* \\ 1, & x > x^* \end{cases}.$$

Then, y can be written as $y = \beta_0 + \beta_1(x - x^*) + (\beta_2 - \beta_1)(x - x^*)z$.
Let

$$x_1 = x - x^*$$
$$x_2 = (x - x^*)z$$
$$\beta_0{}^* = \beta_0$$
$$\beta_1{}^* = \beta_1$$
$$\beta_2{}^* = \beta_2 - \beta_1$$

Then, $y = \beta_0 * + \beta_1 * x_1 + \beta_2 * x_2$.

(b) If there is a discontinuity at x = x*, then a model that can be used is

$$y = \begin{cases} \beta_0 + \beta_1 x & x \le x^* \\ \alpha_0 + \alpha_1 x & x > x^* \end{cases}$$

Let $$z = \begin{cases} 0, & x \le x^* \\ 1, & x > x^* \end{cases}$$

Then, y can be written as $y = \beta_0 + \beta_1 x + [(\alpha_0 - \beta_0) + (\alpha_1 - \beta_1)x]z = \beta_0^* + \beta_1^* x_1 + \beta_2^* z + \beta_3^* x_2$
where

$$\beta_0^* = \beta_0$$
$$\beta_1^* = \beta_1$$
$$\beta_2^* = \alpha_0 - \beta_0$$
$$\beta_3^* = \alpha_1 - \beta_1$$
$$x_1 = x$$
$$x_2 = xz$$

(c) One could estimate x* as a parameter in the model. A simple approach is to refit the model with different choices for x* and to select the value for x* that minimizes the residual sum of squares.

CHAPTER 13

Section 13-2

13-1 Consider the following computer output.

Source	DF	SS	MS	F	*P*-value
Factor	?	117.4	39.1	?	?
Error	16	396.8	?		
Total	19	514.2			

(a) How many levels of the factor were used in this experiment?
(b) How many replicates did the experimenter use?
(c) Fill in the missing information in the ANOVA table. Use bounds for the *P*-value.
(d) What conclusions can you draw about differences in the factor-level means?

(a) Because factor df = total df – error df = 19-16 = 3 (and the degrees of freedom equals the number of levels minus one), 4 levels of the factor were used.

(b) Because the total df = 19, there were 20 trials in the experiment. Because there are 4 levels for the factor, there were 5 replicates of each level.

(c) From part (a), the factor df = 3
MS(Error) = 396.8/16 = 24.8, f = MS(Factor)/MS(Error) = 39.1/24.8 = 1.58.
From Appendix Table VI, 0.1 < P-value < 0.25

(d) We fail to reject H_0. There are not significance differences in the factor level means at $\alpha = 0.05$.

13-2 Consider the following computer output for an experiment. The factor was tested over four levels.

Source	DF	SS	MS	F	*P*-value
Factor	?	?	330.4716	4.42	?
Error	?	?	?		
Total	31	?			

(a) How many replicates did the experimenter use?
(b) Fill in the missing information in the ANOVA table. Use bounds for the *P*-value.
(c) What conclusions can you draw about differences in the factor-level means?

(a) Because the factor was tested over 4 levels and total degrees of freedom is 31, total number of observations is 31 + 1 = 32. Hence, each level has 32/4 = 8 replicates.

(b) Because the factor was tested over 4 levels there are 3 degrees of freedom for factor. Because there are 31 total degrees of freedom, df(Error) = 28.

Because the MS(Factor)= 330.4716, the SS(Factor) = 3(330.4716) = 991.4148.
Because the F statistic equals MS(Factor)/MS(Error) = 4.42 = 330.4716/MS(Error). Therefore, MS(Error) = 74.76733.

Therefore, SS(Error)/df(Error) = MS(Error) = 74.76733. Therefore, SS(Error) = 28(74.76733) = 2093.485

Therefore, SS(Total) = SS(Factor) + SS(Error) = 2084.900

The P-value corresponds to an F = 4.42 with 3 numerator and 28 denominator degrees of freedom and this equal 0.012.

(c) Because the P-value = 0.012 < 0.05, there are significant differences among the mean levels of the factor at significance level 0.05.

13-3 Consider the following computer output for an experiment.

Source	DF	SS	MS	F	*P*-value
Factor	5	?	?	?	?
Error	?	27.38	?		
Total	29	66.34			

(a) How many replicates did the experimenter use?
(b) Fill in the missing information in the ANOVA table. Use bounds for the *P*-value.
(c) What conclusions can you draw about differences in the factor-level means?
(d) Compute an estimate for σ^2.

(a) Because there are 29 total degrees of freedom there are 30 observations. Because there are 5 degrees of freedom for treatments there are 6 treatments. Therefore, there are 5 replicates for each treatment.

(b) df(Error) = 24, MS(Error) = SS(Error)/df(Error) = 27.38/24 = 1.1408
SS(Treatments) = SS(Total) – SS(Error) = 66.34 – 27.38 = 38.96
MS(Treatments) = SS(Treatments)/df(Treatments) = 38.96/5 = 7.792
F = MS(Treatments)/MS(Error) = 7.792/1.1408
P-value = 0.0004 from software < 0.01

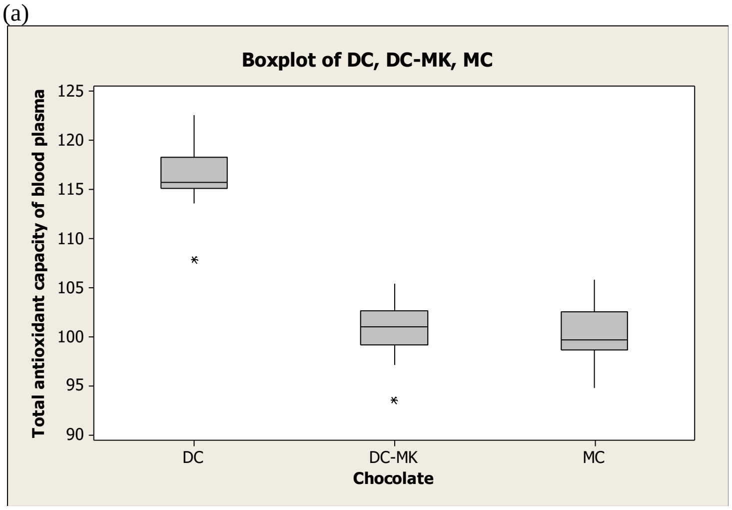
(c) Factor means differ significantly at significance level 0.01

(d) Estimate of σ^2 = MS(Error) = 1.1408

13-4 An article in *Nature* describes an experiment to investigate the effect on consuming chocolate on cardiovascular health ("Plasma Antioxidants from Chocolate," 2003, Vol. 424, pp. 1013). The experiment consisted of using three different types of chocolates: 100 g of dark chocolate, 100 g of dark chocolate with 200 ml of full-fat milk, and 200 g of milk chocolate. Twelve subjects were used, seven women and five men with an average age range of 32.2 ±1 years, an average weight of 65.8 ± 3.1 kg, and body-mass index of 21.9 ± 0.4 kg m^{-2}. On different days, a subject consumed one of the chocolate-factor levels, and one hour later total antioxidant capacity of that person's blood plasma was measured in an assay. Data similar to those summarized in the article follow.

(a) Construct comparative box plots and study the data. What visual impression do you have from examining these plots?
(b) Analyze the experimental data using an ANOVA. If $\alpha = 0.05$, what conclusions would you draw? What would you Conclude if $\alpha = 0.01$?
(c) Is there evidence that the dark chocolate increases the mean antioxidant capacity of the subjects' blood plasma?
(d) Analyze the residuals from this experiment.

(a)



The box plots indicate that the different types of chocolate affect the total antioxidant capacity of blood plasma, especially the dark chocolate.

(b) The computer result is shown below.

One-way ANOVA: DC, DC-MK, MC

```
Source  DF      SS     MS      F      P
Factor   2  1952.6  976.3  93.58  0.000
Error   33   344.3   10.4
Total   35  2296.9

S = 3.230   R-Sq = 85.01%   R-Sq(adj) = 84.10%
```

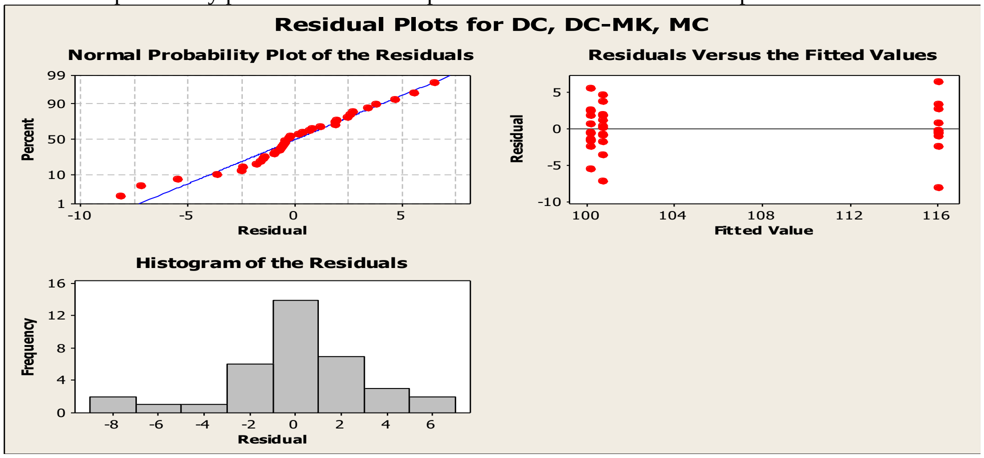
Because the P-value < 0.01 we reject H_0 and conclude that the type of chocolate has an effect on cardiovascular health at $\alpha = 0.05$ or $\alpha = 0.01$.

(c) The computer result is shown below.

```
Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons

Simultaneous confidence level = 88.02%

DC subtracted from:

         Lower   Center    Upper     -+---------+---------+---------+--------
DC+MK  -18.041  -15.358  -12.675      (---*----)
MC     -18.558  -15.875  -13.192     (----*---)
                                     -+---------+---------+---------+--------
                                   -18.0     -12.0      -6.0       0.0

DC+MK subtracted from:

     Lower  Center  Upper     -+---------+---------+---------+--------
MC  -3.200  -0.517  2.166                               (---*----)
                              -+---------+---------+---------+--------
                            -18.0     -12.0      -6.0       0.0
```

The top intervals show differences between the mean antioxidant capacity for DC+MK – DC and MC – DC. Because these intervals are entirely within the negative range (do not include zero) there are significant differences between DC+MK and DC, and MC and DC. This implies that dark chocolate increases the mean antioxidant capacity of the subjects' blood plasma.

(d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



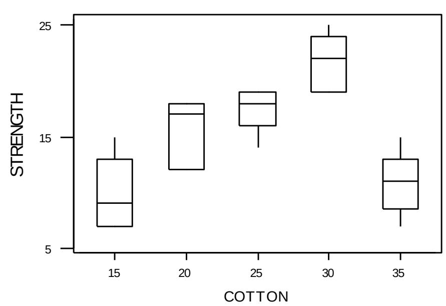
13-5 In *Design and Analysis of Experiments*, 8th edition (John Wiley & Sons, 2012), D. C. Montgomery described an experiment in which the tensile strength of a synthetic fiber was of interest to the manufacturer. It is suspected that strength is related to the percentage of cotton in the fiber. Five levels of cotton percentage were used, and five replicates were run in random order, resulting in the following data.

Cotton Percentage	Observations				
	1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

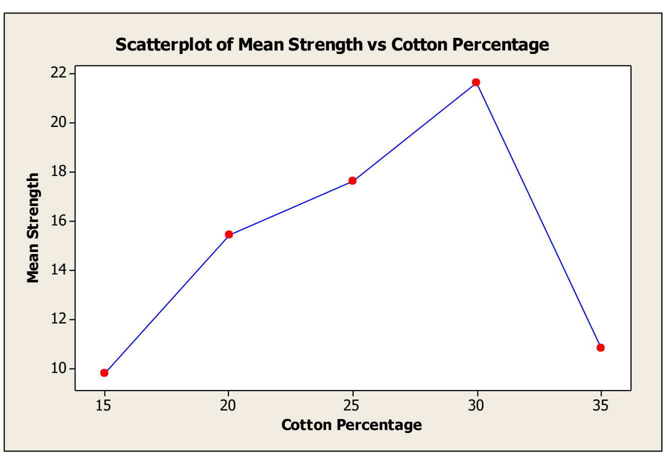
(a) Does cotton percentage affect breaking strength? Draw comparative box plots and perform an analysis of variance. Use $\alpha = 0.05$.
(b) Plot average tensile strength against cotton percentage and interpret the results.
(c) Analyze the residuals and comment on model adequacy.

(a)
```
Analysis of Variance for STRENGTH
Source      DF        SS        MS        F        P
COTTON       4    475.76    118.94    14.76    0.000
Error       20    161.20      8.06
Total       24    636.96
```

Reject H_0 and conclude that cotton percentage affects mean breaking strength.



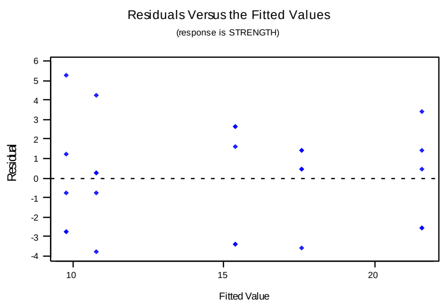
(b) Tensile strength seems to increase up to 30% cotton and declines at 35% cotton.

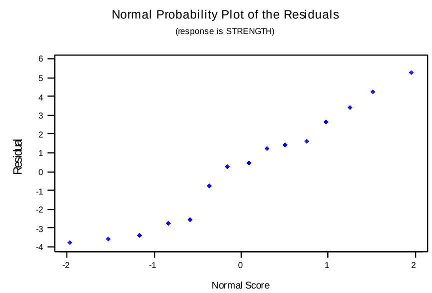


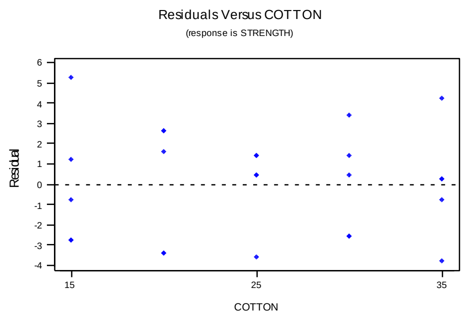

```
                                        Individual 95% CIs For Mean
```

```
                                               Based on Pooled StDev
Level         N       Mean     StDev    ------+---------+---------+---------+
15            5      9.800     3.347    (-----*----)
20            5     15.400     3.130                (----*----)
25            5     17.600     2.074                    (----*----)
30            5     21.600     2.608                            (----*----)
35            5     10.800     2.864      (-----*----)
                                        ------+---------+---------+---------+
Pooled StDev =      2.839                  10.0      15.0      20.0      25.0
```

(c) The normal probability plot and the residual plots show that the model assumptions are reasonable.







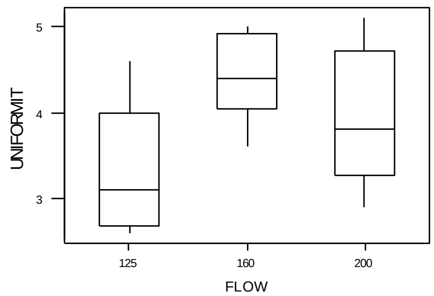
13-6 In "Orthogonal Design for Process Optimization and Its Application to Plasma Etching" (*Solid State Technology*, May 1987), G. Z. Yin and D. W. Jillie described an experiment to determine the effect of C_2F_6 flow rate on the uniformity of the etch on a silicon wafer used in integrated circuit manufacturing. Three flow rates are used in the experiment, and the resulting uniformity (in percent) for six replicates follows.

C_2F_6 Flow (SCCM)	Observations					
	1	2	3	4	5	6
125	2.7	4.6	2.6	3.0	3.2	3.8
160	4.9	4.6	5.0	4.2	3.6	4.2
200	4.6	3.4	2.9	3.5	4.1	5.1

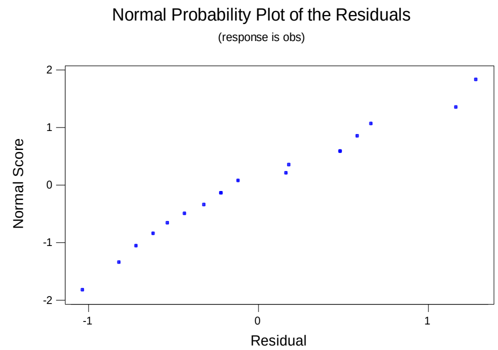
(a) Does C2F6 flow rate affect etch uniformity? Construct box plots to compare the factor levels and perform the analysis of variance. Use $\alpha = 0.05$.
(b) Do the residuals indicate any problems with the underlying assumptions?

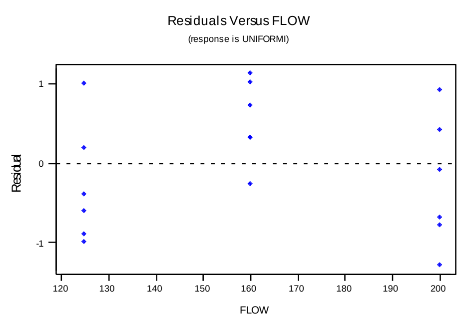
(a)
```
 Analysis of Variance for FLOW
Source       DF        SS         MS       F       P
FLOW          2     3.6478     1.8239    3.59   0.053
Error        15     7.6300     0.5087
Total        17    11.2778
```

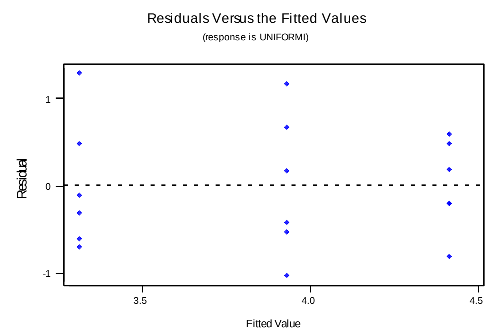
Fail to reject H_0. There is no evidence that flow rate affects etch uniformity.



(b) Residuals are acceptable.







13-7 The compressive strength of concrete is being studied, and four different mixing techniques are being investigated. The following data have been collected.

(a) Test the hypothesis that mixing techniques affect the strength of the concrete. Use $\alpha = 0.05$.
(b) Find the *P*-value for the *F*-statistic computed in part (a).
(c) Analyze the residuals from this experiment.

Mixing Technique	Compressive Strength (psi)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

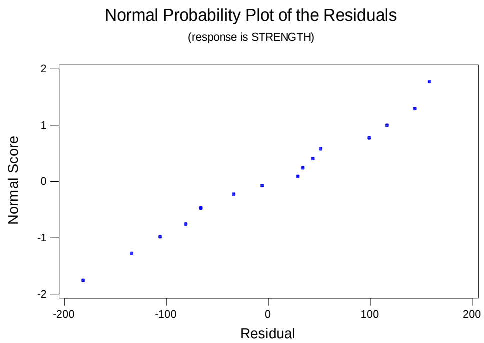
(a)
```
Analysis of Variance for STRENGTH
Source     DF        SS        MS        F        P
TECHNIQU    3    489740    163247    12.73    0.000
```

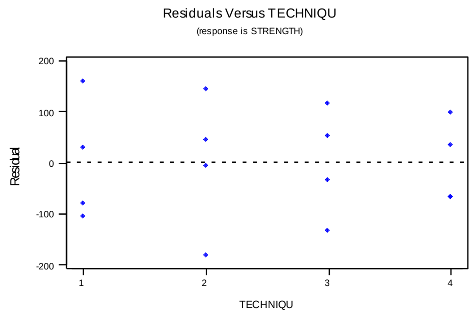
```
Error       12     153908      12826
Total       15     643648
```

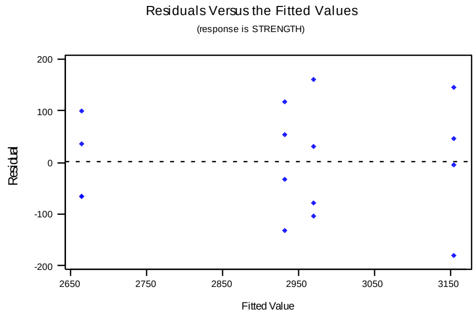
Reject H_0. Techniques affect the mean strength of the concrete.

(b) *P-value* $\cong 0$

(c) Residuals are acceptable







13-8 The response time in milliseconds was determined for three different types of circuits in an electronic calculator. The results are recorded here.

Factor	Subjects (Observations)											
	1	2	3	4	5	6	7	8	9	10	11	12
DC	118.8	122.6	115.6	113.6	119.5	115.9	115.8	115.1	116.9	115.4	115.6	107.9
DC + MK	105.4	101.1	102.7	97.1	101.9	98.9	100.0	99.8	102.6	100.9	104.5	93.5
MC	102.1	105.8	99.6	102.7	98.8	100.9	102.8	98.7	94.7	97.8	99.7	98.6

(a) Using $\alpha = 0.01$, test the hypothesis that the three circuit types have the same response time.
(b) Analyze the residuals from this experiment.
(c) Find a 95% confidence interval on the response time for circuit 3.

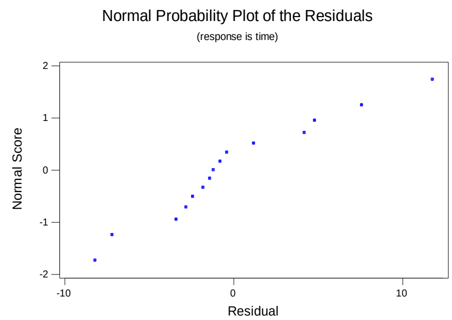
Circuit Type	Response				
1	19	22	20	18	25
2	20	21	33	27	40
3	16	15	18	26	17

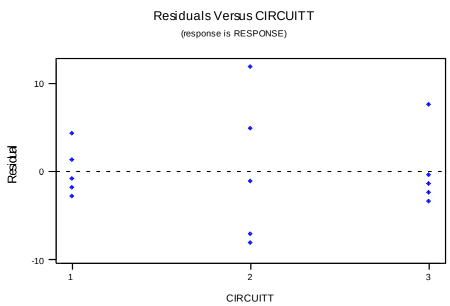
(a) Analysis of Variance for CIRCUIT TYPE

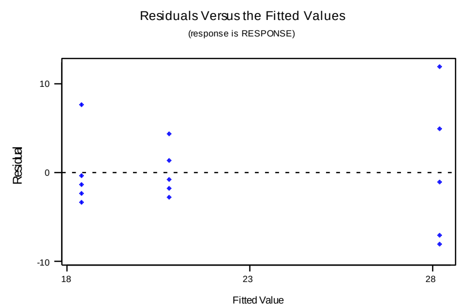
```
Source      DF        SS        MS        F        P
CIRCUITT     2     260.9     130.5     4.01    0.046
Error       12     390.8      32.6
Total       14     651.7
```

Reject H_0

(b)There is some indication of greater variability in circuit two. There is some curvature in the normal probability plot.







(c) 95% Confidence interval on the mean of circuit type 3.

$$\bar{y}_3 - t_{0.025,12}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_3 + t_{0.025,12}\sqrt{\frac{MS_E}{n}}$$

$$18.4 - 2.179\sqrt{\frac{32.6}{5}} \le \mu_1 \le 18.4 + 2.179\sqrt{\frac{32.6}{5}}$$

$$12.84 \le \mu_1 \le 23.96$$

13-9 An electronics engineer is interested in the effect on tube conductivity of five different types of coating for cathode ray tubes in a telecommunications system display device. The following conductivity data are obtained.

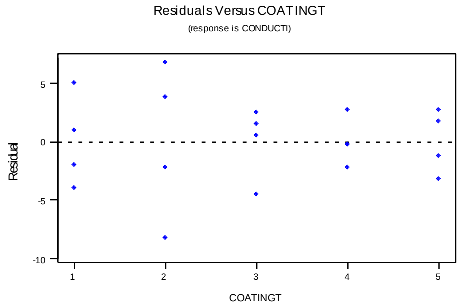
Coating Type	Conductivity			
1	143	141	150	146
2	152	149	137	143
3	134	133	132	127
4	129	127	132	129
5	147	148	144	142

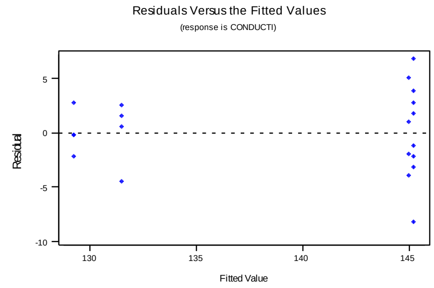
(a) Is there any difference in conductivity due to coating type? Use α = 0.01.
(b) Analyze the residuals from this experiment.
(c) Construct a 95% interval estimate of the coating type 1 mean. Construct a 99% interval estimate of the mean difference between coating types 1 and 4.

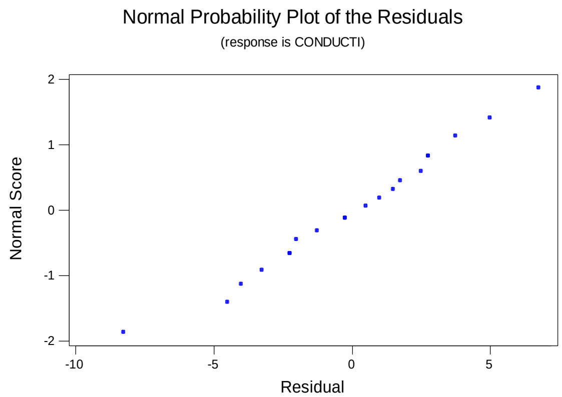
(a)
```
Analysis of Variance for CONDUCTIVITY
Source             DF       SS        MS        F        P
COATINGTYPE         4    1060.5     265.1    16.35    0.000
Error              15     243.3      16.2
Total              19    1303.8
```

Reject H_0, P-value $\cong 0$

(b) There is some indication of that the variability of the response may be increasing as the mean response increases. There appears to be an outlier on the normal probability plot.







(c) 95% Confidence interval on the mean of coating type 1

$$\bar{y}_1 - t_{0.025,15}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_1 + t_{0.015,15}\sqrt{\frac{MS_E}{n}}$$

$$145.00 - 2.131\sqrt{\frac{16.2}{4}} \le \mu_1 \le 145.00 + 2.131\sqrt{\frac{16.2}{4}}$$

$$140.71 \le \mu_1 \le 149.29$$

99% confidence interval on the difference between the means of coating types 1 and 4.

$$\bar{y}_1 - \bar{y}_4 - t_{0.005,15}\sqrt{\frac{2MS_E}{n}} \le \mu_1 - \mu_4 \le \bar{y}_1 - \bar{y}_4 + t_{0.005,15}\sqrt{\frac{2MS_E}{n}}$$

$$(145.00 - 129.25) - 2.947\sqrt{\frac{2(16.2)}{4}} \le \mu_1 - \mu_4 \le (145.00 - 129.25) - 2.947\sqrt{\frac{2(16.2)}{4}}$$

$$7.36 \le \mu_1 - \mu_4 \le 24.14$$

13-10 An article in *Environment International* [1992, Vol. 18(4)] described an experiment in which the amount of radon released in showers was investigated. Radon-enriched water was used in the experiment, and six different orifice diameters were tested in shower heads. The data from the experiment are shown in the following table.

Orifice Diameter	Radon Released (%)			
0.37	80	83	83	85
0.51	75	75	79	79
0.71	74	73	76	77
1.02	67	72	74	74
1.40	62	62	67	69
1.99	60	61	64	66

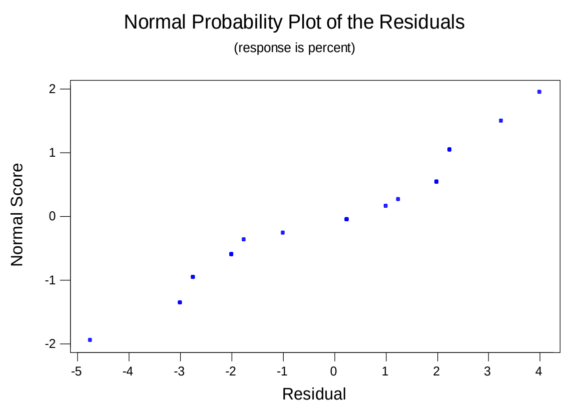
(a) Does the size of the orifice affect the mean percentage of radon released? Use $\alpha = 0.05$.
(b) Find the *P*-value for the *F*-statistic in part (a).
(c) Analyze the residuals from this experiment.
(d) Find a 95% confidence interval on the mean percent of radon released when the orifice diameter is 1.40.

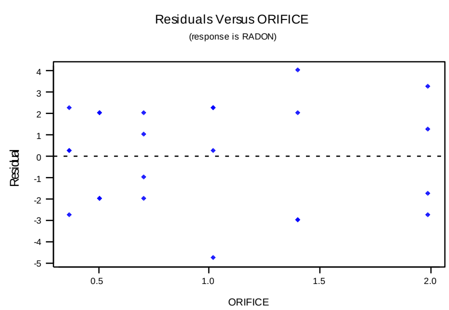
(a)
```
Analysis of Variance for ORIFICE
Source      DF       SS        MS        F        P
ORIFICE      5   1133.37    226.67    30.85    0.000
Error       18    132.25      7.35
Total       23   1265.63
```

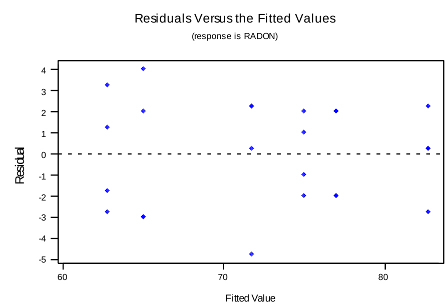
Reject H_0

(b) *P*-value $\cong 0$

(c)







(d) 95% CI on the mean radon released when diameter is 1.40

$$\bar{y}_5 - t_{0.025,18}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_5 + t_{0.015,18}\sqrt{\frac{MS_E}{n}}$$

$$65 - 2.101\sqrt{\frac{7.35}{4}} \le \mu_1 \le 65 + 2.101\sqrt{\frac{7.35}{4}}$$

$$62.15 \le \mu_1 \le 67.84$$

13-11 An article in the *ACI Materials Journal* (1987, Vol. 84, pp. 213–216) described several experiments investigating the rodding of concrete to remove entrapped air. A 3-inch × 6-inch cylinder was used, and the number of times this rod was used is the design variable. The resulting compressive strength of the concrete specimen is the response. The data are shown in the following table.

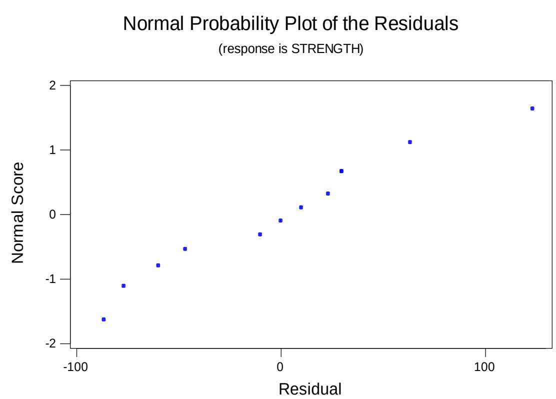
Rodding Level	Compressive Strength		
10	1530	1530	1440
15	1610	1650	1500
20	1560	1730	1530
25	1500	1490	1510

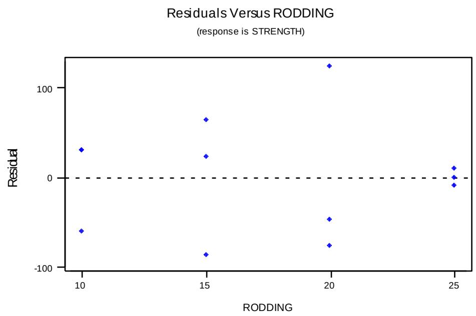
(a) Is there any difference in compressive strength due to the rodding level?
(b) Find the *P*-value for the *F*-statistic in part (a).
(c) Analyze the residuals from this experiment. What conclusions can you draw about the underlying model assumptions?

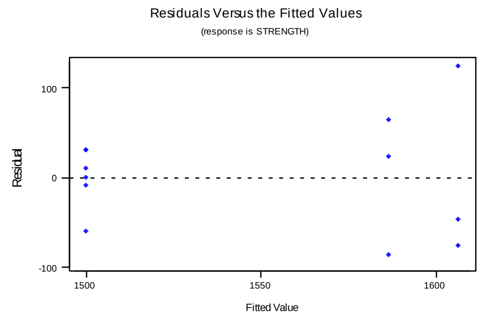
(a)
```
Analysis of Variance for STRENGTH
Source      DF        SS        MS        F        P
RODDING      3     28633      9544     1.87    0.214
Error        8     40933      5117
Total       11     69567
```

Fail to reject H_0

(b) *P-value* = 0.214
(c) The residual plot indicates some concern with nonconstant variance. The normal probability plot looks acceptable.







13-12 An article in the *Materials Research Bulletin* [1991, Vol. 26(11)] investigated four different methods of preparing the superconducting compound $PbMo_6S_8$. The authors contend that the presence of oxygen during the preparation process affects the material's superconducting transition temperature T_c. Preparation methods 1 and 2 use techniques that are designed to eliminate the presence of oxygen, and methods 3 and 4 allow oxygen to be present. Five observations on T_c (in °K) were made for each method, and the results are as follows:

Preparation Method	Transition Temperature T_c(°K)				
1	14.8	14.8	14.7	14.8	14.9
2	14.6	15.0	14.9	14.8	14.7
3	12.7	11.6	12.4	12.7	12.1
4	14.2	14.4	14.4	12.2	11.7

(a) Is there evidence to support the claim that the presence of oxygen during preparation affects the mean transition temperature? Use $\alpha = 0.05$.
(b) What is the *P*-value for the *F*-test in part (a)?
(c) Analyze the residuals from this experiment.
(d) Find a 95% confidence interval on mean T_c when method 1is used to prepare the material.

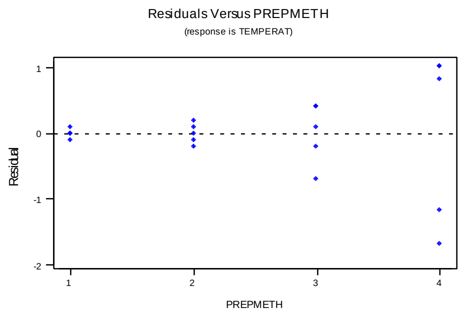
(a) Analysis of Variance of PREPARATION METHOD

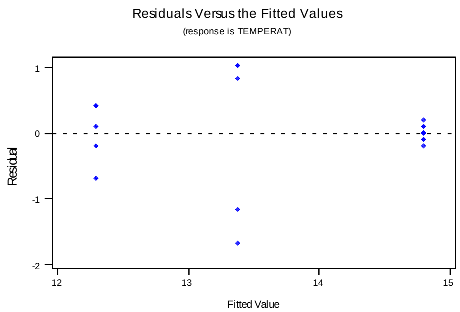
```
Source      DF       SS        MS        F        P
PREPMETH     3    22.124     7.375    14.85    0.000
Error       16     7.948     0.497
Total       19    30.072
```

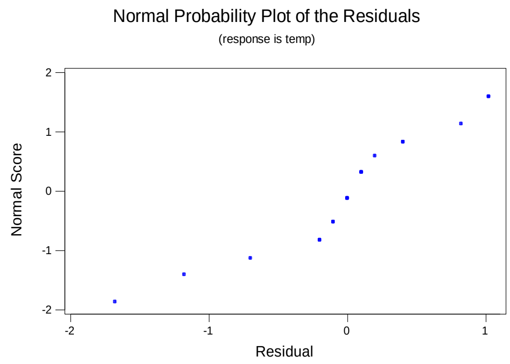
Reject H_0

(b) *P*-value ≅ 0

(c) There are some differences in the amount variability at the different preparation methods and there is some curvature in the normal probability plot. There are also some potential problems with the constant variance assumption apparent in the fitted value plot.







(d) 95% Confidence interval on the mean of temperature for preparation method 1

$$\bar{y}_1 - t_{0.025,16}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_1 + t_{0.015,16}\sqrt{\frac{MS_E}{n}}$$

$$14.8 - 2.120\sqrt{\frac{0.497}{5}} \le \mu_3 \le 14.8 + 2.120\sqrt{\frac{0.497}{5}}$$

$$14.13 \le \mu_1 \le 15.47$$

13-13 A paper in the *Journal of the Association of Asphalt Paving Technologists* (1990, Vol. 59) described an experiment to determine the effect of air voids on percentage retained strength of asphalt. For purposes of the experiment, air voids are controlled at three levels; low (2–4%), medium (4–6%), and high (6–8%). The data are shown in the following table.

Air Voids	Retained Strength (%)							
Low	106	90	103	90	79	88	92	95
Medium	80	69	94	91	70	83	87	83
High	78	80	62	69	76	85	69	85

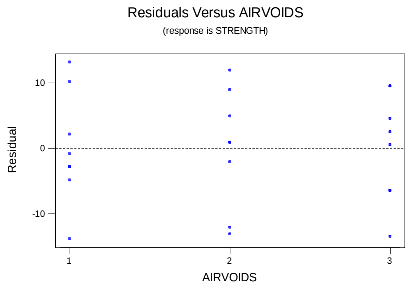
(a) Do the different levels of air voids significantly affect mean retained strength? Use α = 0.01.
(b) Find the *P*-value for the *F*-statistic in part (a).
(c) Analyze the residuals from this experiment.
(d) Find a 95% confidence interval on mean retained strength where there is a high level of air voids.
(e) Find a 95% confidence interval on the difference in mean retained strength at the low and high levels of air voids.

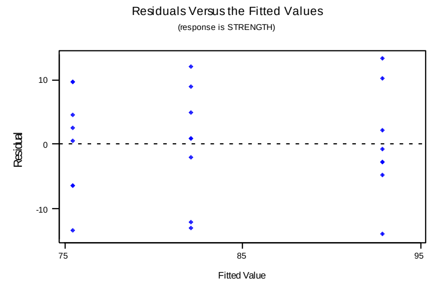
(a)
```
Analysis of Variance for STRENGTH
Source      DF        SS        MS        F        P
AIRVOIDS     2    1230.3     615.1     8.30    0.002
Error       21    1555.8      74.1
Total       23    2786.0
```

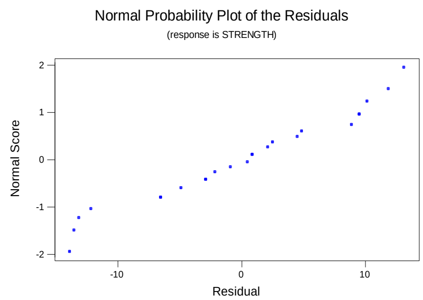
Reject H_0

(b) P-value = 0.002

(c) The residual plots indicate that the constant variance assumption is reasonable. The normal probability plot has some curvature in the tails but appears reasonable.







(d) 95% Confidence interval on the mean of retained strength where there is a high level of air voids

$$\bar{y}_3 - t_{0.025,21}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_3 + t_{0.015,21}\sqrt{\frac{MS_E}{n}}$$

$$75.5 - 2.080\sqrt{\frac{74.1}{8}} \le \mu_3 \le 75.5 + 2.080\sqrt{\frac{74.1}{8}}$$

$$69.17 \le \mu_1 \le 81.83$$

(e) 95% confidence interval on the difference between the means of retained strength at the high level and the low levels of air voids.

$$\bar{y}_1 - \bar{y}_3 - t_{0.025,21}\sqrt{\frac{2MS_E}{n}} \le \mu_1 - \mu_3 \le \bar{y}_1 - \bar{y}_3 + t_{0.025,21}\sqrt{\frac{2MS_E}{n}}$$

$$(92.875 - 75.5) - 2.080\sqrt{\frac{2(74.1)}{8}} \le \mu_1 - \mu_4 \le (92.875 - 75.5) + 2.080\sqrt{\frac{2(74.1)}{8}}$$

$$8.42 \le \mu_1 - \mu_4 \le 26.33$$

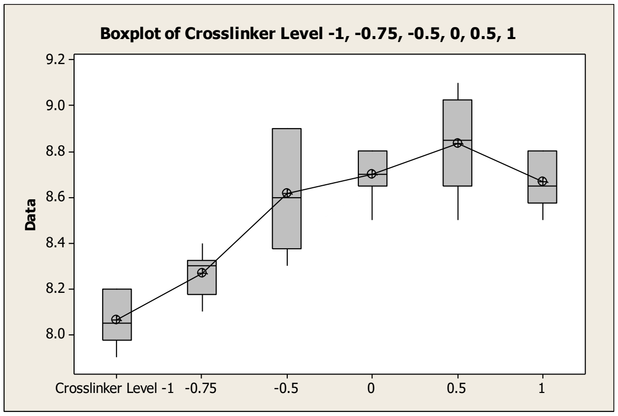
13-14 An article in *Quality Engineering* ["Estimating Sources of Variation: A Case Study from Polyurethane Product Research" (1999–2000, Vol. 12, pp. 89–96)] reported a study on the effects of additives on final polymer properties. In this case, polyurethane additives were referred to as cross-linkers. The average domain spacing was the measurement of the polymer property. The data are as follows:

Cross-Linker Level	Domain Spacing (nm)					
−1	8.2	8	8.2	7.9	8.1	8
−0.75	8.3	8.4	8.3	8.2	8.3	8.1
−0.5	8.9	8.7	8.9	8.4	8.3	8.5
0	8.5	8.7	8.7	8.7	8.8	8.8
0.5	8.8	9.1	9.0	8.7	8.9	8.5
1	8.6	8.5	8.6	8.7	8.8	8.8

(a) Is there a difference in the cross-linker level? Draw comparative box plots and perform an analysis of variance. Use $\alpha = 0.05$.
(b) Find the *P*-value of the test. Estimate the variability due to random error.
(c) Plot average domain spacing against cross-linker level and interpret the results.
(d) Analyze the residuals from this experiment and comment on model adequacy.

(a)
ANOVA

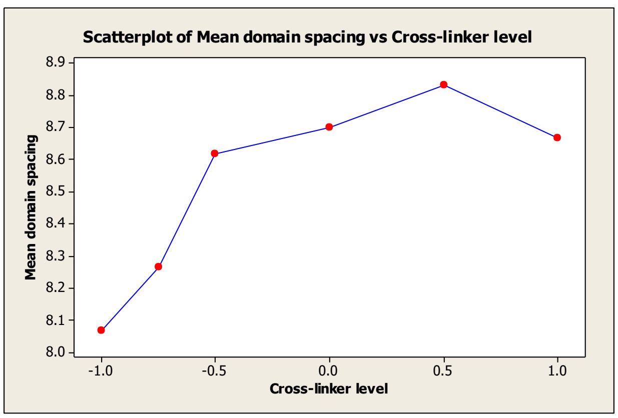
Source	DF	SS	MS	F	P
Factor	5	2.5858	0.5172	18.88	0.000
Error	30	0.8217	0.0274		
Total	35	3.4075			



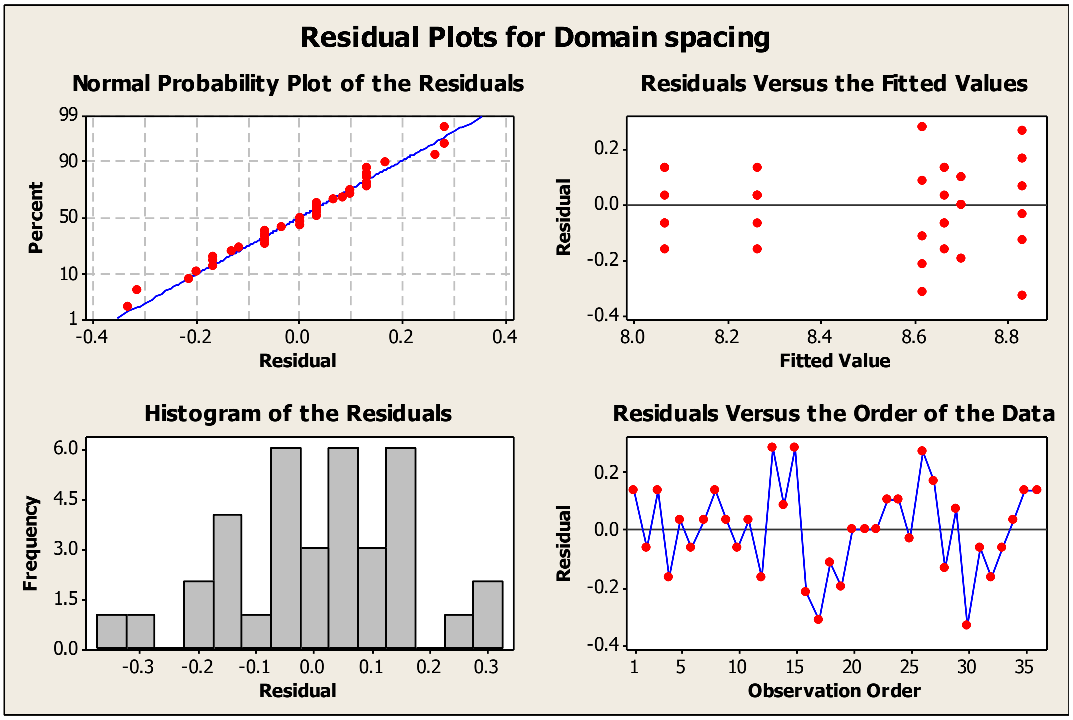
Yes, the box plot and ANOVA show that there is a difference in the cross-linker level.

(b) Anova table in part (a) showed the p-value = 0.000 < α = 0.05. Therefore there is at least one level of cross-linker is different. The variability due to random error is $SS_E = 0.8217$

(c) Domain spacing seems to increase up to the 0.5 cross-linker level and declines once cross-linker level reaches 1.



(d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



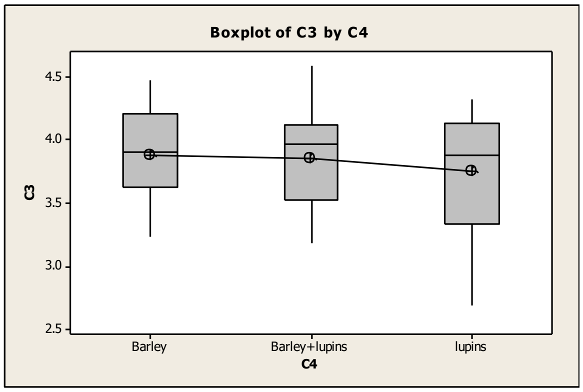
13-15 In the book *Analysis of Longitudinal Data*, 2nd ed., (2002, Oxford University Press), by Diggle, Heagerty, Liang, and Zeger, the authors analyzed the effects of three diets on the protein content of cow's milk. The data shown here were collected after one week and include 25 cows on the barley diet and 27 cows each on the other two diets:

Diet	Protein Content of Cow's Milk													
Barley	3.63	3.24	3.98	3.66	4.34	4.36	4.17	4.4	3.4	3.75	4.2	4.02	4.02	3.9
Barley + lupins	3.38	3.8	4.17	4.59	4.07	4.32	3.56	3.67	4.15	3.51	4.2	4.12	3.52	4.08
Lupins	3.69	4.2	3.31	3.13	3.73	4.32	3.04	3.84	3.98	4.18	4.2	4.1	3.25	3.34
Diet (continued)														
Barley	3.81	3.62	3.66	4.44	4.23	3.82	3.53	4.47	3.93	3.27	3.3			
Barley + lupins	4.02	3.18	4.11	3.27	3.27	3.97	3.31	4.12	3.92	3.78	4	4.37	3.79	
Lupins	3.5	4.13	3.21	3.9	3.5	4.1	2.69	4.3	4.06	3.88	4	3.67	4.27	

(a) Does diet affect the protein content of cow's milk? Draw comparative box plots and perform an analysis of variance. Use $\alpha = 0.05$.
(b) Find the *P*-value of the test. Estimate the variability due to random error.
(c) Plot average protein content against diets and interpret the results.
(d) Analyze the residuals and comment on model adequacy.

(a) No, the diet does not affect the protein content of cow's milk.

Comparative boxplots

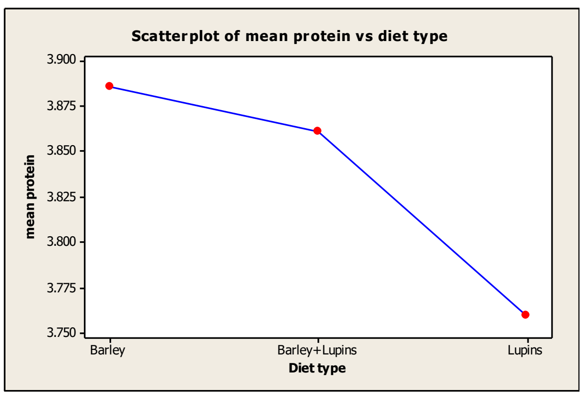


ANOVA

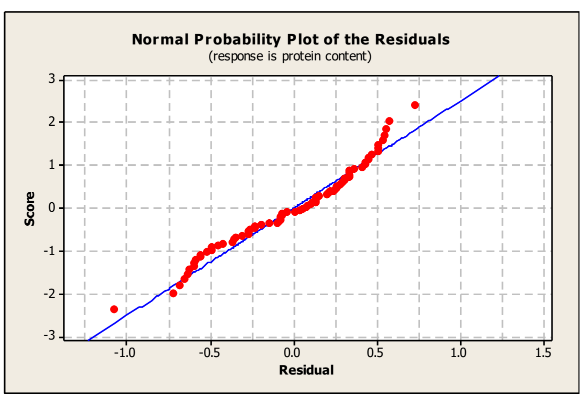
```
Source  DF      SS     MS     F      P
C4       2   0.235  0.118  0.72  0.489
Error   76  12.364  0.163
Total   78  12.599

       S = 0.4033   R-Sq = 1.87%   R-Sq(adj) = 0.00%
```

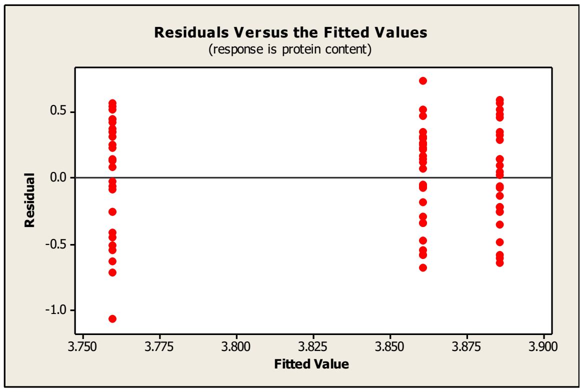
(b) P-value = 0.489. The variability due to random error is $SS_E = 0.146$.

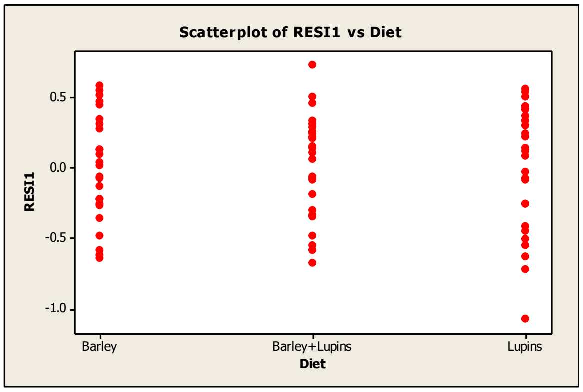
(c) The Barley diet has the highest average protein content and lupins the lowest.



(d) Based on the residual plots, no violation of the ANOVA assumptions is detected.







13-16 An article in *Journal of Food Science* [2001, Vol. 66(3), pp. 472–477] reported on a study of potato spoilage based on different conditions of acidified oxine (AO), which is a mixture of chlorite and chlorine dioxide. The data follow:

AO Solution (ppm)	% Spoilage		
50	100	50	60
100	60	30	30
200	60	50	29
400	25	30	15

(a) Do the AO solutions differ in the spoilage percentage? Use $\alpha = 0.05$.
(b) Find the *P*-value of the test. Estimate the variability due to random error.
(c) Plot average spoilage against AO solution and interpret the results. Which AO solution would you recommend for use in practice?
(d) Analyze the residuals from this experiment.

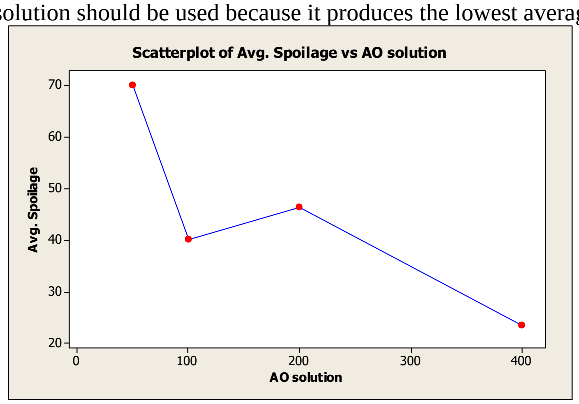
(a) From the analysis of variance shown below, $F_{0.05,3,8} = 4.07 > F_0 = 3.43$ there is no difference in the spoilage percentage when using different AO solutions.

ANOVA

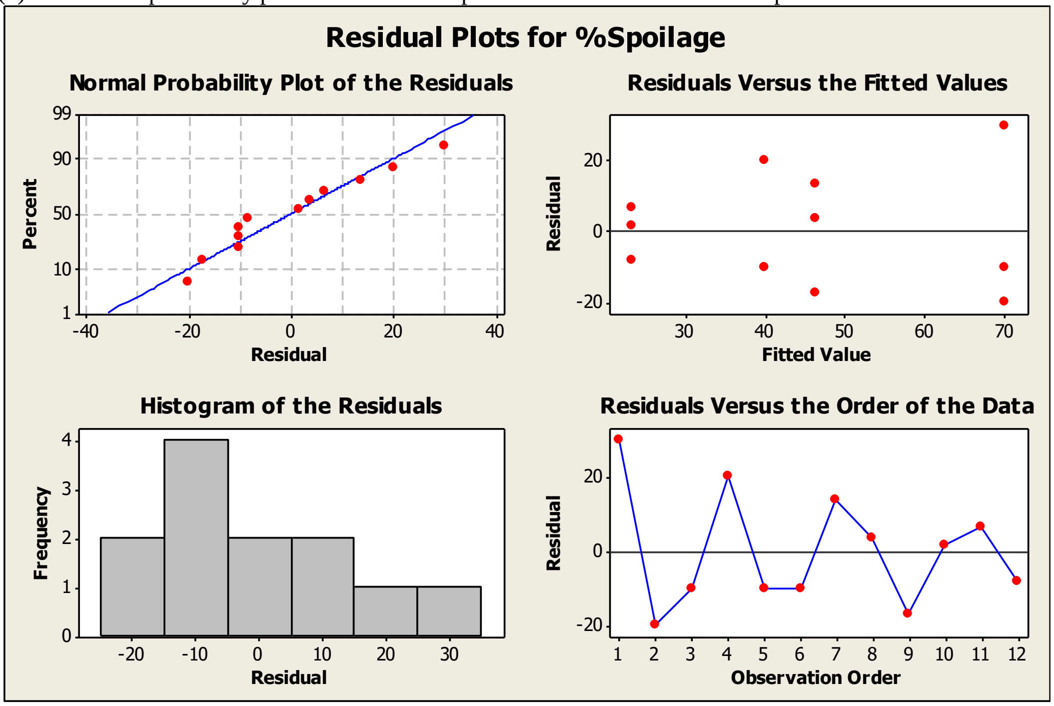
```
Source         DF    SS    MS     F      P
AO solutions    3  3364  1121  3.43  0.073
Error           8  2617   327
Total          11  5981
```

(b) From the table above, the P-value = 0.073 and the variability due to random error is $SS_E = 2617$.

(c) A 400ppm AO solution should be used because it produces the lowest average spoilage percentage.



(d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



13-17 An experiment was run to determine whether four specific firing temperatures affect the density of a certain type of brick. The experiment led to the following data.

Temperature (°F)	Density						
100	21.8	21.9	21.7	21.6	21.7	21.5	21.8
125	21.7	21.4	21.5	21.5	—	—	—
150	21.9	21.8	21.8	21.6	21.5	—	—
175	21.9	21.7	21.8	21.7	21.6	21.8	—

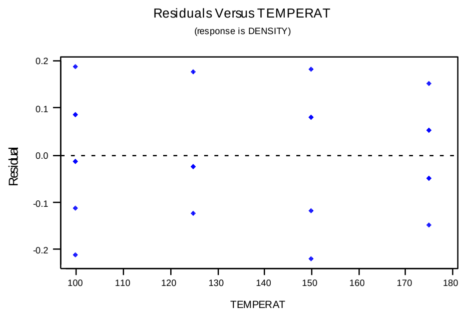
(a) Does the firing temperature affect the density of the bricks? Use $\alpha = 0.05$.
(b) Find the *P*-value for the *F*-statistic computed in part (a).
(c) Analyze the residuals from the experiment.

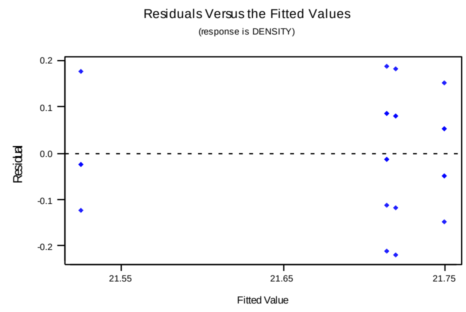
(a) Analysis of Variance for TEMPERATURE

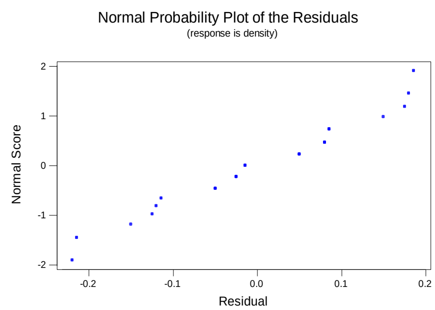
```
Source      DF       SS        MS        F        P
TEMPERAT     3     0.1391    0.0464     2.62    0.083
Error       18     0.3191    0.0177
Total       21     0.4582
```

Fail to reject H_0

(b) *P*-value = 0.083
(c) Residuals are acceptable.





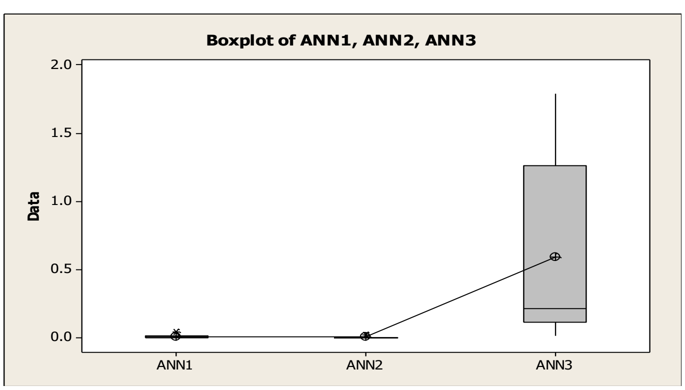


3-18 An article in *Scientia Iranica* ["Tuning the Parameters of an Artificial Neural Network (ANN) Using Central Composite Design and Genetic Algorithm" (2011, Vol. 18(6), pp. 1600–608)], described a series of experiments to tune parameters in artificial neural networks. One experiment considered the relationship between model fitness [measured by the square root of mean square error (RMSE) on a separate test set of data] and model complexity that were controlled by the number of nodes in the two hidden layers. The following data table (extracted from a much larger data set) contains three different ANNs: ANN1 has 33 nodes in layer 1 and 30 nodes in layer 2, ANN2 has 49 nodes in layer 1 and 45 nodes in layer 2, and ANN3 has 17 nodes in layer 1 and 15 nodes in layer 2.

ANN type	RMSE							
ANN1	0.0121	0.0132	0.0011	0.0023	0.0391	0.0054	0.0003	0.0014
ANN2	0.0031	0.0006	0	0	0.022	0.0019	0.0007	0
ANN3	0.1562	0.2227	0.0953	0.8911	1.3892	0.0154	1.7916	0.1992

(a) Construct a box plot to compare the different ANNs.
(b) Perform the analysis of variance with α = 0.05. What is the *P*-value?
(c) Analyze the residuals from the experiment.
(d) Calculate a 95% confidence interval on RMSE for ANN2.

(a)



One-way ANOVA: ANN1, ANN2, ANN3

```
Source  DF     SS     MS     F      P
Factor   2  1.848  0.924  6.02  0.009
Error   21  3.225  0.154
Total   23  5.073

S = 0.3919   R-Sq = 36.43%   R-Sq(adj) = 30.37%


                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level  N    Mean   StDev  ---------+---------+---------+---------+
ANN1   8  0.0094  0.0130  (--------*---------)
ANN2   8  0.0035  0.0075  (--------*---------)
ANN3   8  0.5951  0.6786                     (---------*--------)
                          ---------+---------+---------+---------+
                                0.00      0.30      0.60      0.90

Pooled StDev = 0.3919
```

From the ANOVA table. the P-value = 0.009.

(b) Plots of residual follow. There is substantially lower variability in the measurements when the mean is low. This is evident in the plot of residuals versus the fitted values.

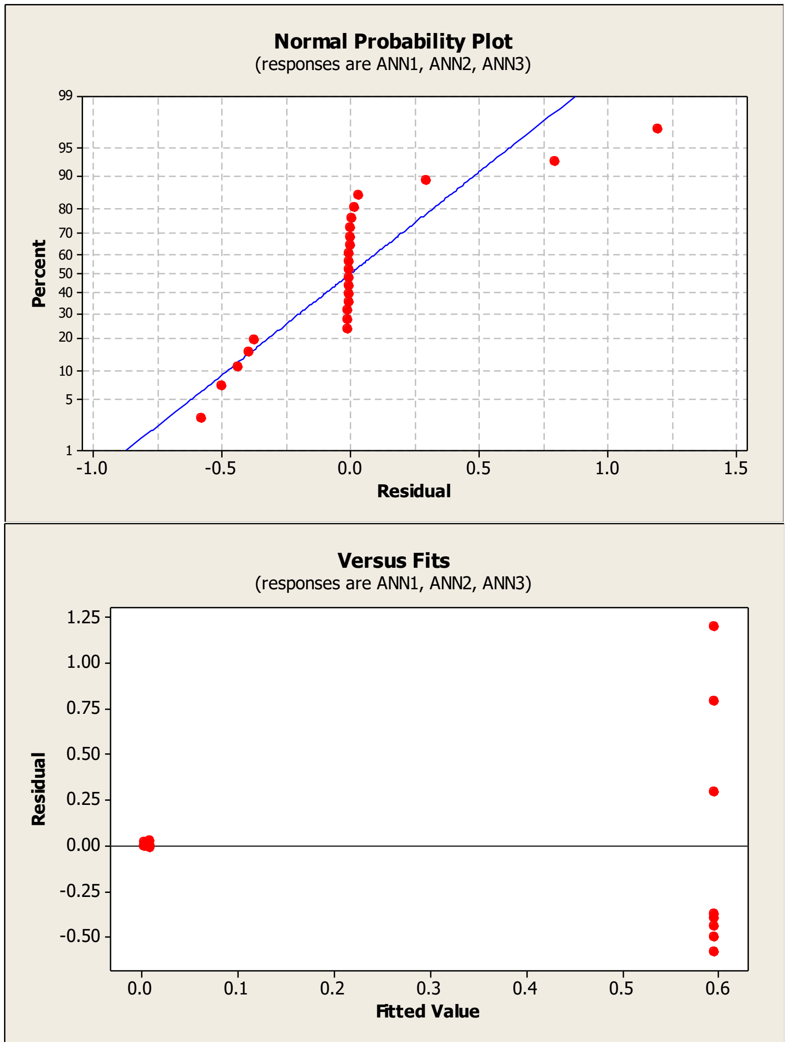
(c) A confidence interval for the mean for ANN2 is obtained as follows.
$t_{0.05/2,\,21}$ = 2.0796, and the pooled standard deviation is 0.3919

$$0.0035 - 2.0796 \times \frac{0.3919}{\sqrt{8}} < \mu_2 < 0.0035 + 2.0796 \times \frac{0.3919}{\sqrt{8}}$$
$$-0.285 < \mu_2 < 0.292$$

However, because the variance differs between the ANNs, one might use only the data from ANN2 to construct a confidence interval. Then

$t_{0.05/2,\,7}$ = 2.3646 and the standard deviation is 0.0075

$$0.0035 - 2.3646 \times \frac{0.0075}{\sqrt{8}} < \mu_2 < 0.0035 + 2.2.3646 \times \frac{0.0075}{\sqrt{8}}$$
$$-0.00277 < \mu_2 < 0.00977$$

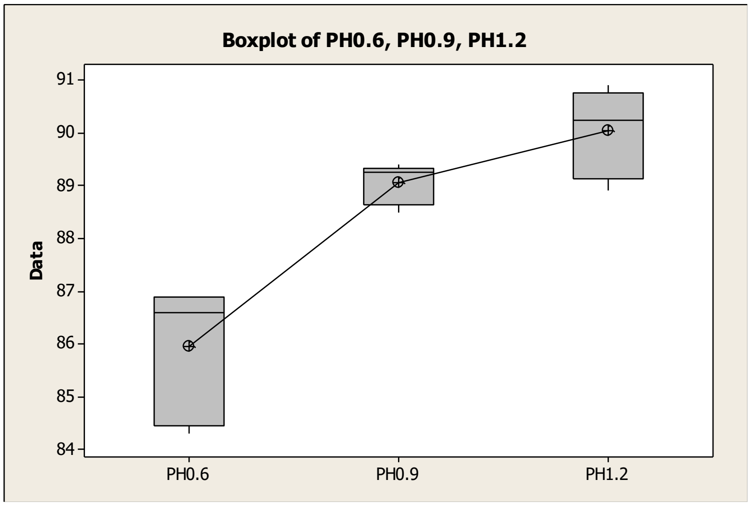


13-19 An article in *Fuel Processing Technology* ("Application of the Factorial Design of Experiments to Biodiesel Production from Lard," 2009, Vol. 90, pp. 1447–1451) described an experiment to investigate the effect of potassium hydroxide in synthesis of biodiesel. It is suspected that potassium hydroxide (PH) is related to fatty acid methyl esters (FAME) which are key elements in biodiesel. Three levels of PH concentration were used, and six replicates were run in a random order. Data are shown in the following table.

PH concentration (wt. %)	FAME concentration (wt. %)					
0.6	84.3	84.5	86.5	86.7	86.9	86.9
0.9	89.3	89.4	88.5	88.7	89.2	89.3
1.2	90.2	90.3	88.9	89.2	90.7	90.9

(a) Construct box plots to compare the factor levels.
(b) Construct the analysis of variance. Are there any differences in PH concentrations at $\alpha = 0.05$? Calculate the *P*-value.
(c) Analyze the residuals from the experiment.
(d) Plot average FAME against PH concentration and interpret your results.
(e) Compute a 95% confidence interval on mean FAME when the PH concentration is 1.2.

(a)



(b)

Anova: Single Factor

SUMMARY

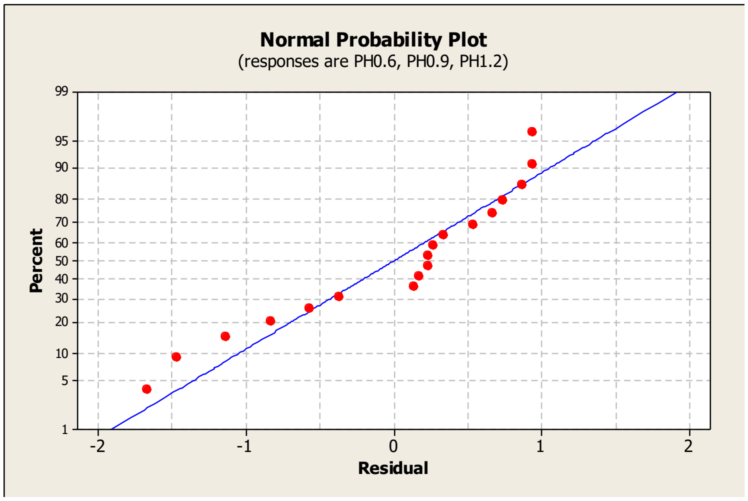
Groups	*Count*	*Sum*	*Average*	*Variance*
0.6	6	515.8	85.96667	1.498667
0.9	6	534.4	89.06667	0.138667
1.2	6	540.2	90.03333	0.654667

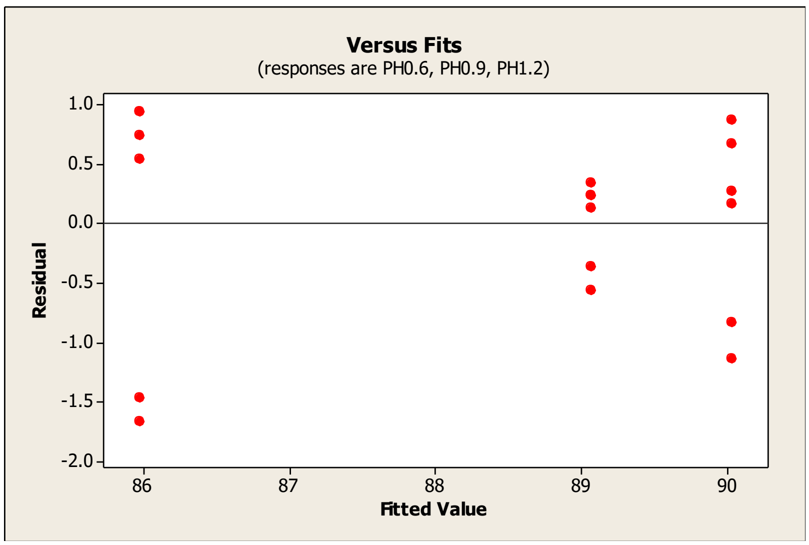
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	54.16444	2	27.08222	35.44793	2.07E-06	3.68232
Within Groups	11.46	15	0.764			
Total	65.62444	17				

There are significant differences between the groups. The P=value = 2.07E-6.

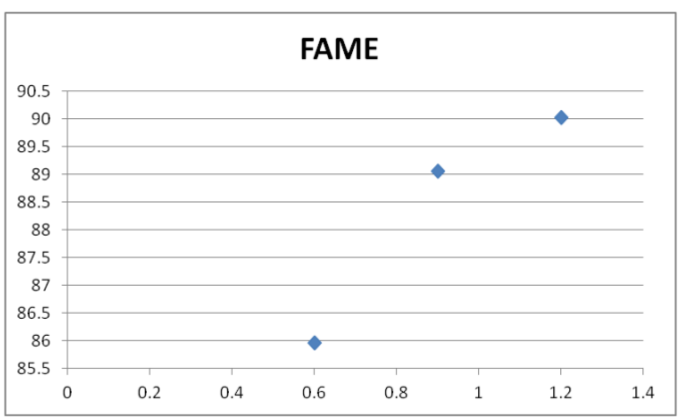
(c)





There is some concern with the normal probability plot, but it does not indicate a serious departure from assumptions. There is also some difference in variation between the groups.

(d) The mean of FAME tends to increase with PH.



(e) $t_{0.05/2,\,15} = 2.1315$ and the pooled estimate of variance from MS(Error) is 0.764. The 95% confidence interval is

$90.03333 \pm 2.1315(0.764/6)^{1/2} = (89.273, 90.794)$

13-20 For each of the following exercises, use the previous data to complete these parts.

(a) Apply Fisher's LSD method with α = 0.05 and determine which levels of the factor differ.
(b) Use the graphical method to compare means described in this section and compare your conclusions to those from Fisher's LSD method.

Chocolate type in Exercise 13-4. Use α = 0.05.

(a)

```
Source      DF      SS     MS      F      P
Chocolate    2  1952.6  976.3  93.58  0.000
Error       33   344.3   10.4
Total       35  2296.9

S = 3.230   R-Sq = 85.01%   R-Sq(adj) = 84.10%


                          Individual 95% CIs For Mean Based on
                          Pooled StDev
Level   N    Mean  StDev   ---+---------+---------+---------+------
DC     12  116.06   3.53                                  (---*---)
DC+MK  12  100.70   3.24     (--*---)
MC     12  100.18   2.89    (--*---)
                           ---+---------+---------+---------+------
                          100.0     105.0     110.0     115.0

Pooled StDev = 3.23


Grouping Information Using Fisher Method

Chocolate   N     Mean  Grouping
DC         12  116.058  A
DC+MK      12  100.700    B
MC         12  100.183    B

Means that do not share a letter are significantly different.


Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Chocolate

Simultaneous confidence level = 88.02%


Chocolate = DC subtracted from:

Chocolate    Lower   Center    Upper
DC+MK      -18.041  -15.358  -12.675
MC         -18.558  -15.875  -13.192

Chocolate     -+---------+---------+---------+--------
DC+MK          (---*----)
MC            (----*---)
              -+---------+---------+---------+--------
           -18.0     -12.0      -6.0       0.0


Chocolate = DC+MK subtracted from:
```

```
Chocolate   Lower  Center  Upper     -+---------+---------+---------+--------
MC         -3.200  -0.517  2.166                                 (---*----)
                                     -+---------+---------+---------+--------
                                  -18.0     -12.0      -6.0       0.0
```

(b) The standard error of a mean is $3.230/12^{1/2} = 0.932$. From the graphical method, group DC is significantly different from the others and this agrees with Fisher's method.

13-21 Cotton percentage in Exercise 13-5. Use $\alpha = 0.05$.

```
Fisher's pairwise comparisons
Family error rate = 0.264
Individual error rate = 0.0500
Critical value = 2.086
Intervals for (column level mean) - (row level mean)
                 15          20          25          30
      20      -9.346
              -1.854
      25     -11.546      -5.946
              -4.054       1.546
      30     -15.546      -9.946      -7.746
              -8.054      -2.454      -0.254
      35      -4.746       0.854       3.054       7.054
               2.746       8.346      10.546      14.546
```

Significant differences are detected between levels 15 and 20, 15 and 25, 15 and 30, 20 and 30, 20 and 35, 25 and 30, 25 and 35, and 30 and 35.

13-22 Flow rate in Exercise 13-6. Use $\alpha = 0.01$.

```
Fisher's pairwise comparisons
Family error rate = 0.117
Individual error rate = 0.0500
Critical value = 2.131

Intervals for (column level mean) - (row level mean)
          125       160
160    -1.9775
       -0.2225
250    -1.4942     -0.3942
       0.2608       1.3608
```

There are significant differences between levels 125 and 160.

13-23 Mixing technique in Exercise 13-7. Use $\alpha = 0.05$.

```
Fisher's pairwise comparisons
Family error rate = 0.184
Individual error rate = 0.0500
Critical value = 2.179
Intervals for (column level mean) - (row level mean)
                  1           2           3
       2        -360
                 -11
       3        -137          48
                 212         397
       4         130         316          93
                 479         664         442
```

Significance differences between levels 1 and 2, 1 and 4, 2 and 3, 2 and 4, and 3 and 4.

13-24 Circuit type in Exercise 13-8. Use $\alpha = 0.01$.

```
Fisher's pairwise comparisons

Family error rate = 0.0251
Individual error rate = 0.0100

Critical value = 3.055

Intervals for (column level mean) - (row level mean)

              1           2

 2      -18.426
          3.626

 3       -8.626      -1.226
         13.426      20.826
```

No significant differences at α = 0.01.

13-25 Coating type in Exercise 13-9. Use α = 0.01.

```
Fisher's pairwise comparisons

Family error rate = 0.0649
Individual error rate = 0.0100
Critical value = 2.947
Intervals for (column level mean) - (row level mean)

              1           2           3           4

 2       -8.642
          8.142
 3        5.108       5.358
         21.892      22.142
 4        7.358       7.608      -6.142
         24.142      24.392      10.642
 5       -8.642      -8.392     -22.142     -24.392
          8.142       8.392      -5.358      -7.608
```

Significant differences between 1 and 3, 1 and 4, 2 and 3, 2 and 4, 3 and 5, 4 and 5.

13-26 Preparation method in Exercise 13-12. Use α = 0.05.

```
Fisher's pairwise comparisons
 Family error rate = 0.189
Individual error rate = 0.0500
Critical value = 2.120
Intervals for (column level mean) - (row level mean)
                   1           2           3
      2     -0.9450
             0.9450
      3      1.5550      1.5550
             3.4450      3.4450
      4      0.4750      0.4750     -2.0250
             2.3650      2.3650     -0.1350
```

There are significant differences between levels 1 and 3, 4; 2 and 3, 4; and 3 and 4.

13-27 Air voids in Exercise 13-13. Use α = 0.05.

```
Fisher's pairwise comparisons
Family error rate = 0.118
Individual error rate = 0.0500
```

```
Critical value = 2.080
Intervals for (column level mean) - (row level mean)
                   1           2
       2       1.799
              19.701
       3       8.424      -2.326
              26.326      15.576
```

Significant differences between levels 1 and 2; and 1 and 3.

13-28 Cross-linker Exercise 13-14. Use α = 0.05.

(a) $$LSD = t_{0.025,25}\sqrt{\frac{2MS_E}{b}} = 2.042\sqrt{\frac{2\times 0.0274}{6}} = 0.1952$$

```
Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Cross-linker level

Simultaneous confidence level = 65.64%

Cross-linker level = -0.5 subtracted from:

Cross-linker
level             Lower   Center    Upper
-0.75           -0.5451  -0.3500  -0.1549
-1              -0.7451  -0.5500  -0.3549
0               -0.1118   0.0833   0.2785
0.5              0.0215   0.2167   0.4118
1               -0.1451   0.0500   0.2451

Cross-linker
level           ---------+---------+---------+---------+
-0.75                  (---*---)
-1                 (---*---)
0                           (---*---)
0.5                           (---*---)
1                          (---*---)
                ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00


Cross-linker level = -0.75 subtracted from:

Cross-linker
level             Lower   Center    Upper
-1              -0.3951  -0.2000  -0.0049
0                0.2382   0.4333   0.6285
0.5              0.3715   0.5667   0.7618
1                0.2049   0.4000   0.5951

Cross-linker
level           ---------+---------+---------+---------+
-1                        (---*---)
0                                    (---*---)
0.5                                    (---*---)
1                                   (---*---)
                ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00


Cross-linker level = -1 subtracted from:
```

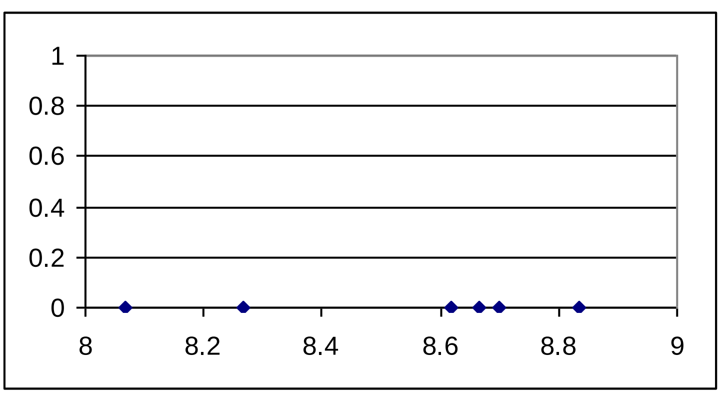
```
Cross-linker
level           Lower  Center   Upper  ---------+---------+---------+---------+
0              0.4382  0.6333  0.8285                            (---*---)
0.5            0.5715  0.7667  0.9618                              (---*---)
1              0.4049  0.6000  0.7951                           (---*---)
                                ---------+---------+---------+---------+
                                      -0.50      0.00      0.50      1.00


Cross-linker level = 0 subtracted from:

Cross-linker
level            Lower   Center   Upper
0.5            -0.0618   0.1333  0.3285
1              -0.2285  -0.0333  0.1618

Cross-linker
level          ---------+---------+---------+---------+
0.5                              (---*---)
1                            (---*---)
               ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00


Cross-linker level = 0.5 subtracted from:

Cross-linker
level            Lower   Center   Upper
1              -0.3618  -0.1667  0.0285

Cross-linker
level          ---------+---------+---------+---------+
1                          (---*---)
               ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00
```

Cross-linker levels -0.5, 0, 0.5 and 1 are not detected to differ. Cross-linker levels -0.75 and -1 are not detected to differ from one other, but both are significantly different to the others.

(b) The mean values are
8.0667, 8.2667, 8.6167, 8.7, 8.8333, 8.6667

$$\hat{\sigma}_{\bar{X}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{0.0274}{6}} = 0.0676$$

The width of a scaled normal distribution is 6(0.0676) = 0.405



With a scaled normal distribution over this plot, the conclusions are similar to those from the LSD method.

13-29 Diets in Exercise 13-15. Use $\alpha = 0.01$.

(a) There is no significant difference in protein content between the three diet types.

```
Fisher 99% Individual Confidence Intervals
All Pairwise Comparisons among Levels of C4

Simultaneous confidence level = 97.33%

C4 = Barley subtracted from:

C4                 Lower   Center   Upper
Barley+lupins  -0.3207  -0.0249  0.2709
lupins         -0.4218  -0.1260  0.1698

C4             -------+---------+---------+---------+--
Barley+lupins      (-----------*-----------)
lupins         (-----------*-----------)
               -------+---------+---------+---------+--
                   -0.25      0.00      0.25      0.50


C4 = Barley+lupins subtracted from:

C4        Lower   Center   Upper  -------+---------+---------+---------+--
lupins  -0.3911  -0.1011  0.1889   (-----------*-----------)
                                  -------+---------+---------+---------+--
                                      -0.25      0.00      0.25      0.50
```

(b) The mean values are: 3.886, 3.8611, 3.76 (barley, b+l, lupins)

```
From the ANOVA the estimate of σ can be obtained
Source  DF      SS     MS     F      P
C4       2   0.235  0.118  0.72  0.489
Error   76  12.364  0.163
Total   78  12.599

S = 0.4033   R-Sq = 1.87%   R-Sq(adj) = 0.00%
The minimum sample size could be used to calculate the standard error of a
sample mean
```

$$\hat{\sigma}_{\overline{X}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{0.163}{25}} = 0.081$$

The graph would not show any differences between the diets.

13-30 Suppose that four normal populations have common variance $\sigma^2 = 25$ and means $\mu_1 = 50$, $\mu_2 = 60$, $\mu_3 = 50$, and $\mu_4 = 60$. How many observations should be taken on each population so that the probability of rejecting the hypothesis of equality of means is at least 0.90? Use $\alpha = 0.05$.

$\overline{\mu} = 55$, $\tau_1 = -5$, $\tau_2 = 5$, $\tau_3 = -5$, $\tau_4 = 5$.

$$\Phi^2 = \frac{n(100)}{4(25)} = n, \qquad a-1=3 \qquad a(n-1) = 4(n-1)$$

Various choices for *n* yield:

n	Φ^2	Φ	a(n-1)	Power=1-β
4	4	2	12	0.80
5	5	2.24	16	0.90

Therefore, $n = 5$ is needed.

13-31 Suppose that five normal populations have common variance $\sigma^2 = 100$ and means $\mu_1 = 175$, $\mu_2 = 190$, $\mu_3 = 160$, $\mu_4 = 200$, and $\mu_5 = 215$. How many observations per population must be taken so that the probability of rejecting the hypothesis of equality of means is at least 0.95? Use $\alpha = 0.01$.

$\bar{\mu} = 188$, $\tau_1 = -13$, $\tau_2 = 2$, $\tau_3 = -28$, $\tau_4 = 12$, $\tau_5 = 27$.

$$\Phi^2 = \frac{n(1830)}{5(100)} = n, \qquad a - 1 = 4 \qquad a(n-1) = 5(n-1)$$

Various choices for n yield:

n	Φ^2	Φ	a(n – 1)	Power =1-β
2	7.32	2.7	5	0.55
3	10.98	3.13	10	0.95

Therefore, $n = 3$ is needed.

13-32 Suppose that four normal populations with common variance σ^2 are to be compared with a sample size of eight observations from each population. Determine the smallest value for $\sum_{i=1}^{4} \tau_i^2 / \sigma^2$ that can be detected with power 90%. Use $\alpha = 0.05$.

From the top chart with $v_1 = 3$ (4 treatments – 1), $v_2 = 28$ (degrees of freedom for error = 4(7) = 28), $\varphi = 2$. Then the smallest value of $\frac{\sum_{i=1}^{4} \tau_i^2}{\sigma^2}$ is $a\varphi^2/n = \varphi^2/2 = 4/2 = 2$, where $a = 4$ and $n = 8$.

13-33 Suppose that five normal populations with common variance σ^2 are to be compared with a sample size of seven observations from each. Suppose that $\tau_1 = \ldots = \tau_4 = 0$. What is the smallest value for τ_5^2/σ^2 that can be detected with power 90% and $\alpha = 0.01$?

From the bottom chart with $v_1 = 4$ (5 treatments – 1), $v_2 = 30$ (degrees of freedom for error = 5(6) = 30), obtain φ is approximately 2.2. Then the smallest value of $\frac{\sum_{i=1}^{5} \tau_i^2}{\sigma^2}$ is $a\varphi^2/n = 5\varphi^2/7 = 3.5$, where $a = 5$ and $n = 7$.

Section 13-3

13-34 An article in the *Journal of the Electrochemical Society* [1992, Vol. 139(2), pp. 524–532)] describes an experiment to investigate the low-pressure vapor deposition of polysilicon. The experiment was carried out in a large-capacity reactor at Sematech in Austin, Texas. The reactor has several wafer positions, and four of these positions were selected at random. The response variable is film thickness uniformity. Three replicates of the experiment were run, and the data are as follows:

(a) Is there a difference in the wafer positions? Use $\alpha = 0.05$.
(b) Estimate the variability due to wafer positions.
(c) Estimate the random error component.
(d) Analyze the residuals from this experiment and comment on model adequacy.

Wafer Position	**Uniformity**		
1	2.76	5.67	4.49
2	1.43	1.70	2.19
3	2.34	1.97	1.47
4	0.94	1.36	1.65

(a)

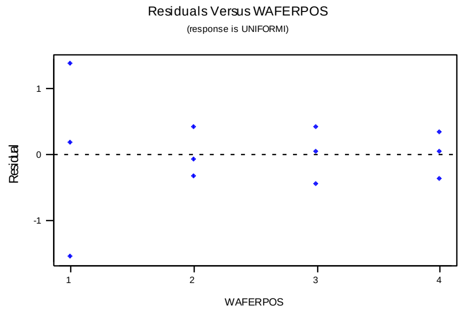
```
Analysis of Variance for UNIFORMITY
Source      DF        SS        MS        F        P
WAFERPOS     3    16.220     5.407     8.29    0.008
Error        8     5.217     0.652
Total       11    21.437
```

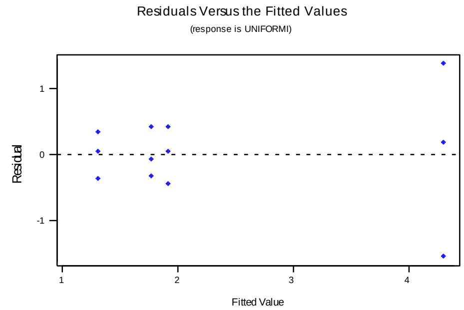
Reject H_0, and conclude that there are significant differences among wafer positions.

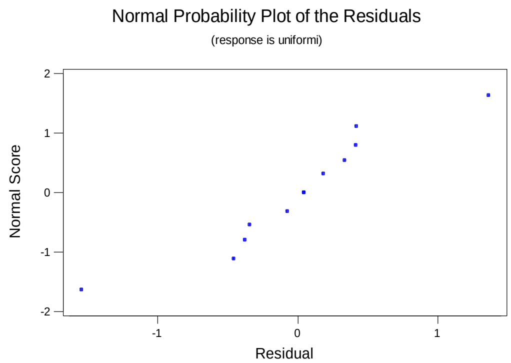
(b) $\hat{\sigma}_{\tau}^{2} = \dfrac{MS_{Treatments} - MS_{E}}{n} = \dfrac{5.407 - 0.652}{3} = 1.585$

(c) $\hat{\sigma}^{2} = MS_{E} = 0.652$

(d) Greater variability at wafer position 1. There is some slight curvature in the normal probability plot.







13-35 A textile mill has a large number of looms. Each loom is supposed to provide the same output of cloth per minute. To investigate this assumption, five looms are chosen at random, and their output is measured at different times. The following data are obtained:

Loom	Output (lb/min)				
1	4.0	4.1	4.2	4.0	4.1
2	3.9	3.8	3.9	4.0	4.0
3	4.1	4.2	4.1	4.0	3.9
4	3.6	3.8	4.0	3.9	3.7
5	3.8	3.6	3.9	3.8	4.0

(a) Are the looms similar in output? Use $\alpha = 0.05$.
(b) Estimate the variability between looms.
(c) Estimate the experimental error variance.
(d) Analyze the residuals from this experiment and check for model adequacy.

(a) Analysis of Variance for OUTPUT

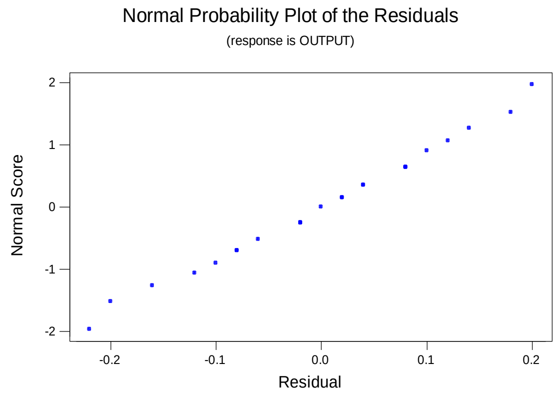
```
Source      DF        SS        MS        F        P
LOOM         4    0.3416    0.0854     5.77    0.003
Error       20    0.2960    0.0148
Total       24    0.6376
```

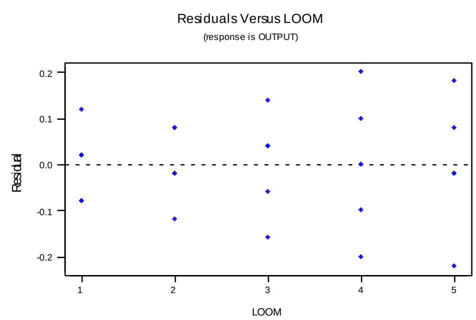
Reject H_0, there are significant differences among the looms.

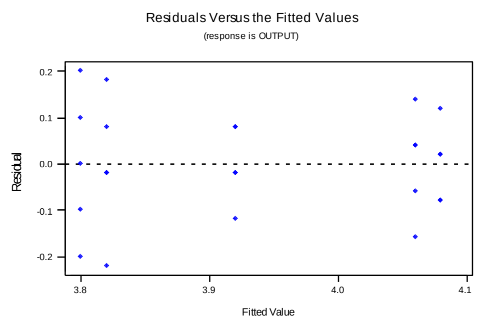
(b) $\hat{\sigma}_{\tau}^{2} = \dfrac{MS_{Treatments} - MS_E}{n} = \dfrac{0.0854 - 0.0148}{5} = 0.01412$

(c) $\hat{\sigma}^2 = MS_E = 0.0148$

(d) Residuals are acceptable.







13-36 In the book *Bayesian Inference in Statistical Analysis* (1973, John Wiley and Sons) by Box and Tiao, the total product yield for five samples was determined randomly selected from each of six randomly chosen batches of raw material.

Batch	Yield (in grams)				
1	1545	1440	1440	1520	1580
2	1540	1555	1490	1560	1495
3	1595	1550	1605	1510	1560
4	1445	1440	1595	1465	1545
5	1595	1630	1515	1635	1625
6	1520	1455	1450	1480	1445

(a) Do the different batches of raw material significantly affect mean yield? Use α = 0.01.
(b) Estimate the variability between batches.
(c) Estimate the variability between samples within batches.
(d) Analyze the residuals from this experiment and check for model adequacy.

(a) Yes, the different batches of raw material significantly affect mean yield at α = 0.01 because the P-value is small.

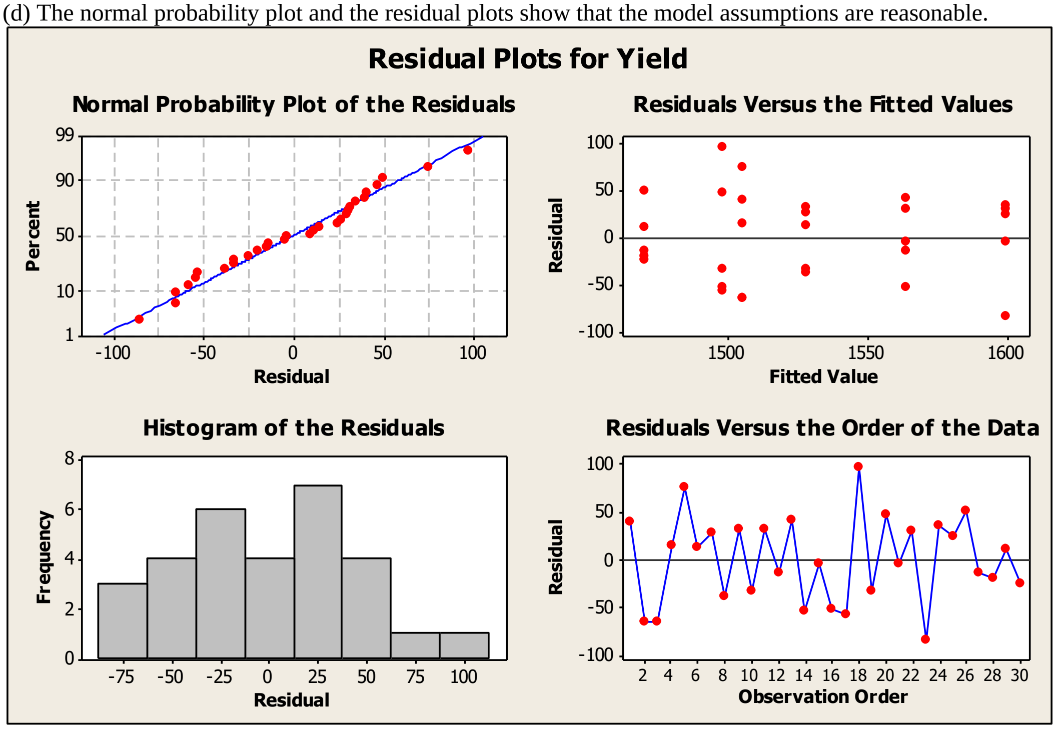
```
Source  DF      SS     MS     F      P
Batch    5   56358  11272  4.60  0.004
Error   24   58830   2451
Total   29  115188
```

(b) Variability between batches

$$\hat{\sigma}_\tau^2 = \frac{MS_{Treatments} - MS_E}{n} = \frac{11272 - 2451}{5} = 1764.2$$

(c) Variability within batches $\hat{\sigma}^2 = MSE = 2451$

(d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



13-37 An article in the *Journal of Quality Technology* [1981, Vol. 13(2), pp. 111–114)] described an experiment that investigated the effects of four bleaching chemicals on pulp brightness. These four chemicals were selected at random from a large population of potential bleaching agents. The data are as follows:

(a) Is there a difference in the chemical types? Use α = 0.05.
(b) Estimate the variability due to chemical types.
(c) Estimate the variability due to random error.
(d) Analyze the residuals from this experiment and comment on model adequacy.

Chemical	Pulp Brightness				
1	77.199	74.466	92.746	76.208	82.876
2	80.522	79.306	81.914	80.346	73.385
3	79.417	78.017	91.596	80.802	80.626
4	78.001	78.358	77.544	77.364	77.386

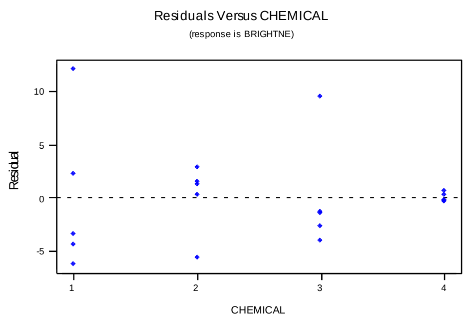
(a)
```
Analysis of Variance for BRIGHTNENESS
Source     DF      SS       MS       F       P
CHEMICAL    3     54.0     18.0    0.75    0.538
Error      16    384.0     24.0
Total      19    438.0
```

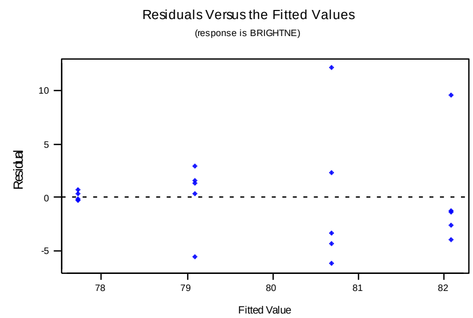
Fail to reject H_0, there is no significant difference among the chemical types.

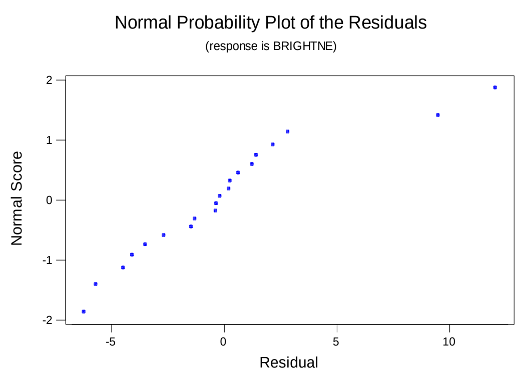
(b) $\hat{\sigma}_\tau^2 = \dfrac{18.0 - 24.0}{5} = -1.2$ set equal to 0

(c) $\hat{\sigma}^2 = 24.0$

(d) Variability is smaller in chemical 4. There is some curvature in the normal probability plot.







13-38 Consider the vapor-deposition experiment described in Exercise 13-34.

(a) Estimate the total variability in the uniformity response.
(b) How much of the total variability in the uniformity response is due to the difference between positions in the reactor?
(c) To what level could the variability in the uniformity response be reduced if the position-to-position variability in the Reactor could be eliminated? Do you believe this is a substantial reduction?

(a) $\hat{\sigma}^2_{total} = \hat{\sigma}^2_{position} + \hat{\sigma}^2 = 2.237$

(b) $\dfrac{\hat{\sigma}^2_{position}}{\hat{\sigma}^2_{total}} = 0.709$

(c) It could be reduced to 0.6522. This is a reduction of approximately 71%.

13-39 Consider the cloth experiment described in Exercise 13-35.

(a) Estimate the total variability in the output response.
(b) How much of the total variability in the output response is due to the difference between looms?
(c) To what level could the variability in the output response be reduced if the loom-to-loom variability could be eliminated? Do you believe this is a significant reduction?

From 13-35 we have

$$\hat{\sigma}_\tau^2 = \frac{MS_{Treatments} - MS_E}{n} = \frac{0.0854 - 0.0148}{5} = 0.01412$$

$$\hat{\sigma}^2 = MS_E = 0.0148$$

(a) The total variability is 0.0141 + 0.0148 = 0.0289.
(b)The proportion due to the difference in looms is 0.01412/0.0289 = 48.8%
(c) Variability could be reduced to 0.0148 if loom differences are eliminated. This is a substantial reduction of approximately one half.

13-40 Reconsider Exercise 13-8 in which the effect of different circuits on the response time was investigated. Suppose that the three circuits were selected at random from a large number of circuits.

(a) How does this change the interpretation of the experiment?
(b) What is an appropriate statistical model for this experiment?
(c) Estimate the parameters of this model.

(a) Now this is a random effects experiment.

(b) $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ where τ_i and ε_{ij} are random variables.

(c) We obtain the following ANOVA table.

```
Source  DF     SS     MS     F      P
Factor   2  260.9  130.5  4.01  0.046
Error   12  390.8   32.6
Total   14  651.7

S = 5.707   R-Sq = 40.04%   R-Sq(adj) = 30.04%
```

$$\hat{\sigma}_\tau^2 = \frac{MS_{Treatments} - MS_E}{n} = \frac{130.5 - 32.6}{5} = 19.58$$

$$\hat{\sigma}^2 = MS_E = 32.6$$

13-41 Reconsider Exercise 13-15 in which the effect of different diets on the protein content of cow's milk was investigated. Suppose that the three diets reported were selected at random from a large number of diets. To simplify, delete the last two observations in the diets with n = 27 (to make equal sample sizes).

(a) How does this change the interpretation of the experiment?
(b) What is an appropriate statistical model for this experiment?
(c) Estimate the parameters of this model.

(a) Instead of testing the hypothesis that the individual treatment effects are zero, we are testing whether there is variability in protein content between all diets.

$$H_0 : \sigma_\tau^2 = 0$$

$$H_1 : \sigma_\tau^2 \neq 0$$

(b) The statistical model is

$$y = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}$$

$$\varepsilon_i \sim N(0,\sigma^2) \text{ and } \tau_i \sim N(0,\sigma_\tau^2)$$

(c) The **last TWO observations were omitted** from two diets to generate equal sample sizes with n = 25.

ANOVA: Protein versus DietType

```
Analysis of Variance for Protein

Source    DF      SS      MS     F      P
DietType   2  0.2689  0.1345  0.82  0.445
```

```
Error      72  11.8169  0.1641
Total      74  12.0858


S = 0.405122   R-Sq = 2.23%   R-Sq(adj) = 0.00%
```

$$\sigma^2 = MS_E = 0.1641$$

$$\sigma_\tau^2 = \frac{MS_{tr} - MS_E}{n} = \frac{0.1345 - 0.1641}{25} = -0.001184$$

Section 13-4

13-42 Consider the following computer output from a RCBD.

Source	DF	SS	MS	F	P
Factor	?	193.800	64.600	?	?
Block	3	464.218	154.739		
Error	?	?	4.464		
Total	15	698.190			

(a) How many levels of the factor were used in this experiment?
(b) How many blocks were used in this experiment?
(c) Fill in the missing information. Use bounds for the *P*-value.
(d) What conclusions would you draw if α = 0.05? What would you conclude if α = 0.01?

(a) $MS_{factor} = \frac{SS_{factor}}{DF_{factor}}$, $DF_{factor} = \frac{SS_{factor}}{MS_{factor}} = \frac{193.8}{64.6} = 3$

The levels of the factor = DF for the factor + 1 = 3 + 1 = 4. Therefore, 4 levels of the factor are used in this experiment. This can also be obtained from the result that the error degrees of freedom equal the product of the degrees of freedom for factor and block.

Let dfF and dfB denote the degrees of freedom for factors and blocks, respectively. Therefore, the total degrees of freedom = 15 = dfT + dfB + (dfT)(dfB) = dfT + 3 + 3(dfT). Therefore dfT = 3.

(b) Because the number of blocks = DF of block +1 = 3 + 1 =4. There are 4 blocks used in this experiment.

(c) From part a), DF factor = 3.

$$F = \frac{MS_{factor}}{MS_{error}} = \frac{64.6}{4.464} = 14.4713$$

P-value < 0.0

$DF_{error} = DF_{Total} - DF_{Factor} - DF_{Block} = 15 - 3 - 3 = 9$.

$MS_{block} = \frac{SS_{block}}{DF_{block}}$, $SS_{block} = MS_{block}\, DF_{block} = 4.464(9) = 40.176$

(d) Because the P-value < 0.01, we reject H_0. There are significance differences in the factor level means at $\alpha = 0.05$ or $\alpha = 0.01$.

13-43 Consider the following computer output from a RCBD. There are four levels of the factor and five blocks.

Source	DF	SS	MS	F	P
Factor	?	?	115.2067	3.49809	?
Block	?	?	71.9775		
Error	?	?	?		
Total	?	?			

(a) Fill in the missing information. Use bounds for the *P*-value.
(b) What conclusions would you draw if $\alpha = 0.05$? What would you conclude if $\alpha = 0.01$?

(a) Because there are 4 levels and 5 blocks, there are 20 trials. Therefore, df(Total)= 19 and df(Factor) = 3 and df(Block) = 4. Therefore, df(Error) = df(Total) – df(Factor) – df(Block) = 19 – 3 – 4 = 12.
Also, SS(Factor) = MS(Factor)(df(Factor)) = 115.2067(3) = 345.620
Also, SS(Block) = MS(Block)(df(Block)) = 71.9775(4) = 345.620

Because F = 3.49809 = MS(Factor)/MS(Error), MS(Error) = 115.2067/3.49809 = 32.934
Therefore, SS(Error) = MS(Error) (df(Error)) = 32.934(12) = 395.210

Because F = 3.49809 with 3 numerator and 12 denominator degrees of freedom, the *P*-value = 0.0497.

(b) Because the 0.01 < *P*-value < 0.05, reject H_0 at $\alpha = 0.05$, but fail to reject H0 at $\alpha = 0.01$.

13-44 Exercise 13-4 introduced you to an experiment to investigate the potential effect of consuming chocolate on cardiovascular health. The experiment was conducted as a completely randomized design, and the exercise asked you to use the ANOVA to analyze the data and draw conclusions. Now assume that the experiment had been conducted as an RCBD with the subjects considered as blocks. Analyze the data using this assumption. What conclusions would you draw (using $\alpha = 0.05$) about the effect of the different types of chocolate on cardiovascular health? Would your conclusions change if $\alpha = 0.01$?

The output from computer software follows.

```
Source  DF       SS       MS       F      P
Factor   2  1952.64  976.322  147.35  0.000
Block   11   198.54   18.049    2.72  0.022
Error   22   145.77    6.626
Total   35  2296.95

S = 2.574   R-Sq = 93.65%   R-Sq(adj) = 89.90%
```

Because the *P*-value for the factor is near zero, there are significant differences in the factor level means at $\alpha = 0.05$ or $\alpha = 0.01$.

13-45 Reconsider the experiment of Exercise 13-5. Suppose that the experiment was conducted as a RCBD with blocks formed by days (denoted as columns in the data table). In the experiment, the primary interest is still in the effect of cotton percentage and day is considered a nuisance factor.

(a) Consider day as a block, and re-estimate the ANOVA.
(b) Does cotton percentage still affect strength at $\alpha = 0.05$?
(c) Compare the conclusions here with those obtained from the analysis without blocks.

(a)
```
Source  DF       SS       MS      F      P
Cotton   2  161.733  80.8667  11.78  0.004
Block    4   46.267  11.5667   1.68  0.246
Error    8   54.933   6.8667
Total   14  262.933

S = 2.620   R-Sq = 79.11%   R-Sq(adj) = 63.44%
```

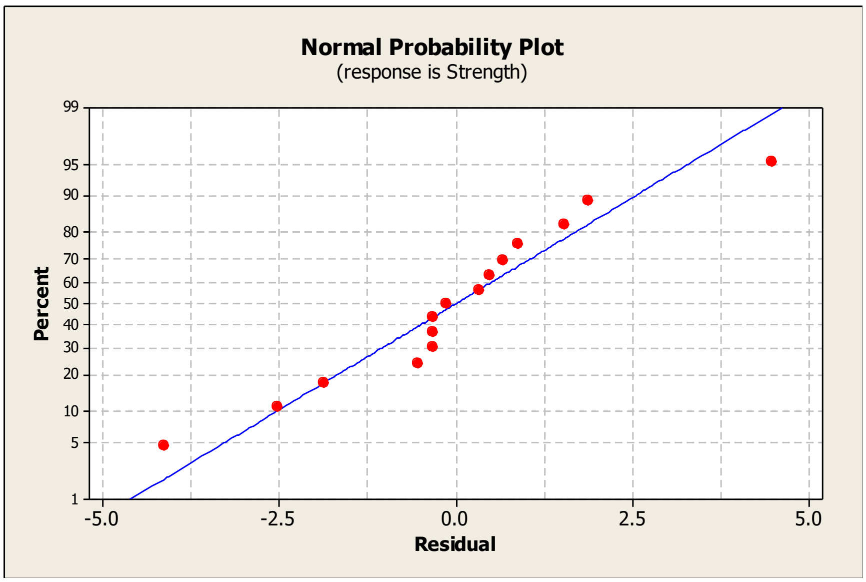
(b) Because the cotton *P*-value = 0.004 < 0.05, the cotton percentage is a significant factor for strength of the fabric.

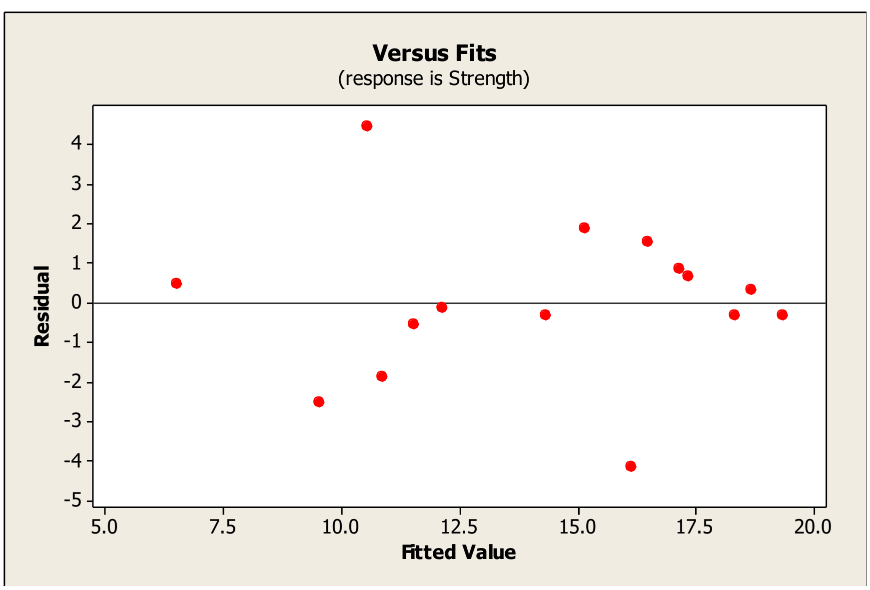
(c) The ANOVA for the analysis without blocks follows. The cotton P-value is still much smaller than 0.05 so the cotton percentage is a significant factor for strength of the fabric.

The MS error is slightly lower than for the blocked analysis. Therefore, blocking is not necessary in an experiment such as this.

```
Source   DF      SS     MS     F      P
Cotton    2  161.73  80.87  9.59  0.003
Error    12  101.20   8.43
Total    14  262.93

S = 2.904   R-Sq = 61.51%   R-Sq(adj) = 55.10%
```

(d) The residual plots follow. There are no serious departures from the assumptions.





13-46 An article in *Quality Engineering* ["Designed Experiment to Stabilize Blood Glucose Levels" (1999–2000, Vol. 12, pp. 83–87)] described an experiment to minimize variations in blood glucose levels. The treatment was the exercise time on a Nordic Track cross-country skier (10 or 20 min). The experiment was blocked for time of day. The data were as follows:

Exercise (min)	Time of Day	Average Blood Glucose
10	pm	71.5
10	am	103.0
20	am	83.5
20	pm	126.0
10	am	125.5
10	pm	129.5
20	pm	95.0
20	am	93.0

(a) Is there an effect of exercise time on the average blood glucose? Use $\alpha = 0.05$.
(b) Find the *P*-value for the test in part (a).
(c) Analyze the residuals from this experiment.

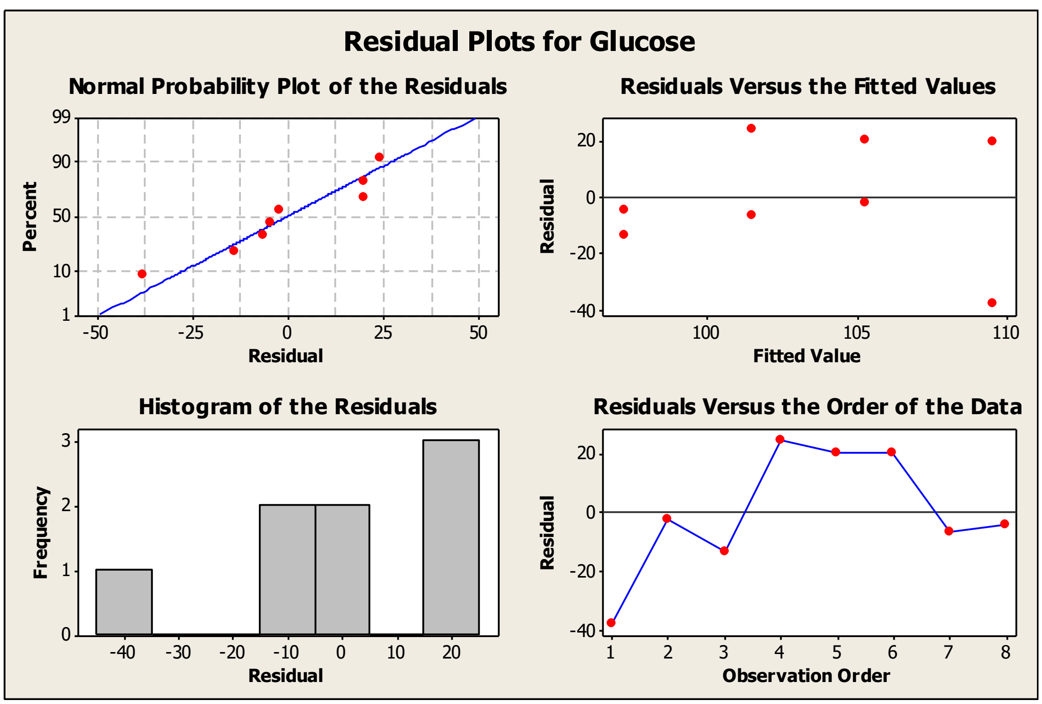
(a) Analysis of variance for Glucose

```
Source   DF       SS       MS     F      P
Time      1    36.13   36.125  0.06  0.819
Min       1   128.00  128.000  0.21  0.669
Error     5  3108.75  621.750
Total     7  3272.88
```

No, there is no effect of exercise time on the average blood glucose.

(b) *P*-value = 0.819

(c) The normal probability plot and the residual plots show that the model assumptions are reasonable.



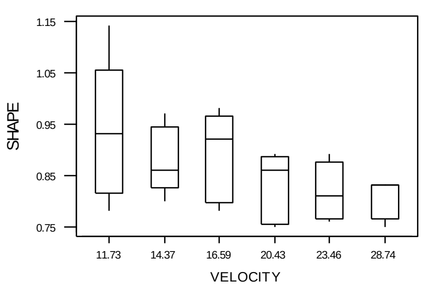
13-47 In "The Effect of Nozzle Design on the Stability and Performance of Turbulent Water Jets" (*Fire Safety Journal*, August 1981, Vol. 4), C. The obald described an experiment in which a shape measurement was determined for several different nozzle types at different levels of jet efflux velocity. Interest in this experiment focuses primarily on nozzle type, and velocity is a nuisance factor. The data are as follows:

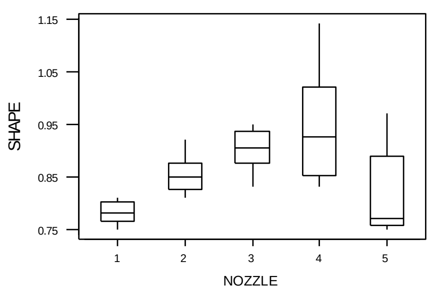
(a) Does nozzle type affect shape measurement? Compare the nozzles with box plots and the analysis of variance.
(b) Use Fisher's LSD method to determine specific differences among the nozzles. Does a graph of the average (or standard deviation) of the shape measurements versus nozzle type assist with the conclusions?
(c) Analyze the residuals from this experiment.

	Jet Efflux Velocity (m/s)					
Nozzle Type	**11.73**	**14.37**	**16.59**	**20.43**	**23.46**	**28.74**
1	0.78	0.80	0.81	0.75	0.77	0.78
2	0.85	0.85	0.92	0.86	0.81	0.83
3	0.93	0.92	0.95	0.89	0.89	0.83
4	1.14	0.97	0.98	0.88	0.86	0.83
5	0.97	0.86	0.78	0.76	0.76	0.75

(a)
```
Analysis of Variance for SHAPE
 Source       DF        SS          MS         F      P
 NOZZLE        4   0.102180    0.025545    8.92  0.000
 VELOCITY      5   0.062867    0.012573    4.39  0.007
 Error        20   0.057300    0.002865
 Total        29   0.222347
```

Reject H_0, nozzle type affects shape measurement.

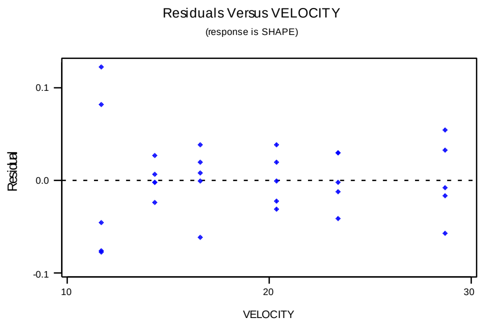


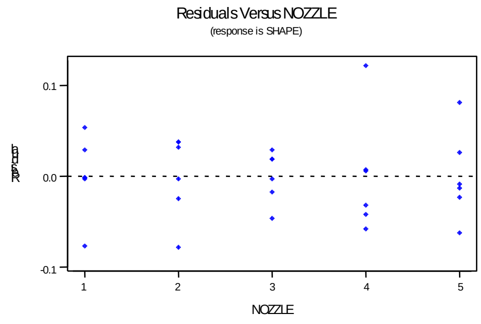


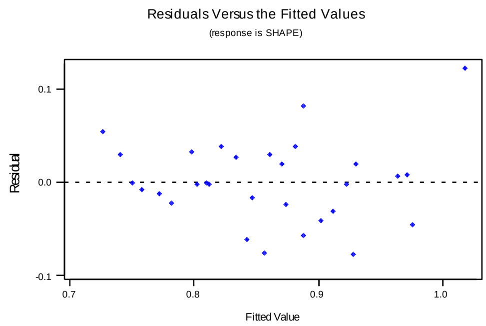
(b)
```
 Fisher's pairwise comparisons
   Family error rate = 0.268
   Individual error rate = 0.0500
   Critical value = 2.060
   Intervals for (column level mean) - (row level mean)
              1           2           3           4
       2   -0.15412
            0.01079
       3   -0.20246    -0.13079
           -0.03754     0.03412
       4   -0.24412    -0.17246    -0.12412
           -0.07921    -0.00754     0.04079
       5   -0.11412    -0.04246     0.00588     0.04754
            0.05079     0.12246     0.17079     0.21246
```

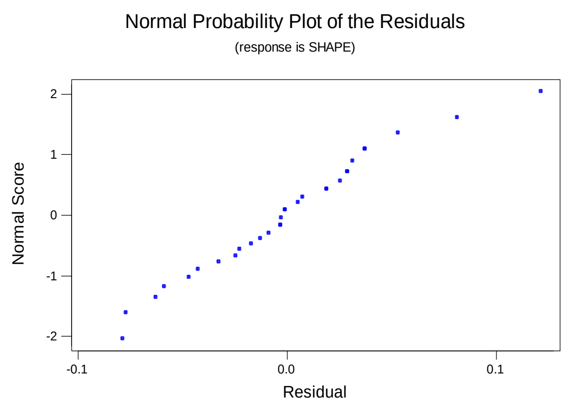
```
There are significant differences between levels 1 and 3, 4; 2 and 4; 3 and 5;
and 4 and 5.
```

(c) The residual analysis shows that there is some inequality of variance. The normal probability plot is acceptable.









13-48 In *Design and Analysis of Experiments*, 8th edition (John Wiley & Sons, 2012), D. C. Montgomery described an experiment that determined the effect of four different types of tips in a hardness tester on the observed hardness of a metal alloy. Four specimens of the alloy were obtained, and each tip was tested once on each specimen, producing the following data:

	Specimen			
Type of Tip	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

(a) Is there any difference in hardness measurements between the tips?
(b) Use Fisher's LSD method to investigate specific differences between the tips.
(c) Analyze the residuals from this experiment.

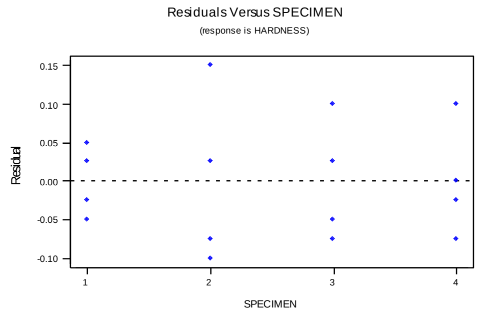
(a) Analysis of Variance of HARDNESS

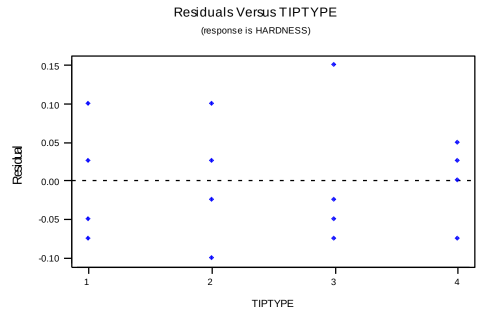
Source	DF	SS	MS	F	P
TIPTYPE	3	0.38500	0.12833	14.44	0.001
SPECIMEN	3	0.82500	0.27500	30.94	0.000
Error	9	0.08000	0.00889		
Total	15	1.29000			

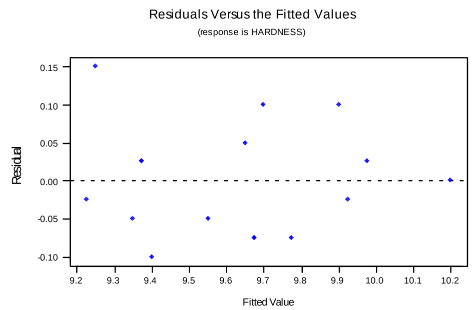
Reject H_0, and conclude that there are significant differences in hardness measurements between the tips.

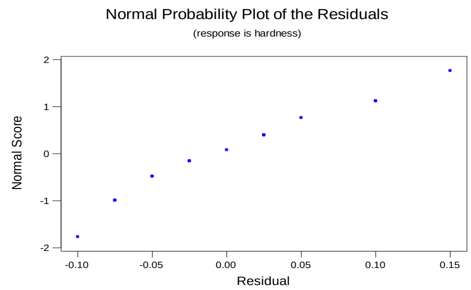
(b)
```
Fisher's pairwise comparisons
Family error rate = 0.184
Individual error rate = 0.0500
Critical value = 2.179
Intervals for (column level mean) - (row level mean)
             1           2           3
2     -0.4481
       0.3981
3     -0.2981     -0.2731
       0.5481      0.5731
4     -0.7231     -0.6981     -0.8481
       0.1231      0.1481     -0.0019

Significant difference between tip types 3 and 4
```

(c) Residuals are acceptable.









13-49 An article in the *American Industrial Hygiene Association Journal* (1976, Vol. 37, pp. 418–422) described a field test for detecting the presence of arsenic in urine samples. The test has been proposed for use among forestry workers because of the increasing use of organic arsenics in that industry. The experiment compared the test as performed by both a trainee and an experienced trainer to an analysis at a remote laboratory. Four subjects were selected for testing and are considered as blocks. The response variable is arsenic content (in ppm) in the subject's urine. The data are as follows:

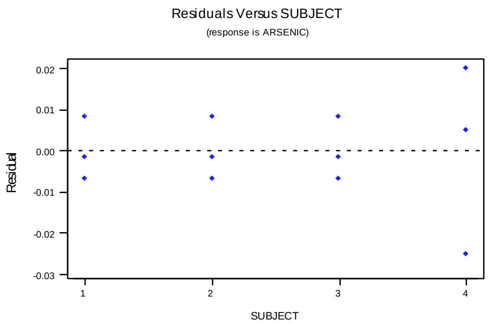
(a) Is there any difference in the arsenic test procedure?
(b) Analyze the residuals from this experiment.

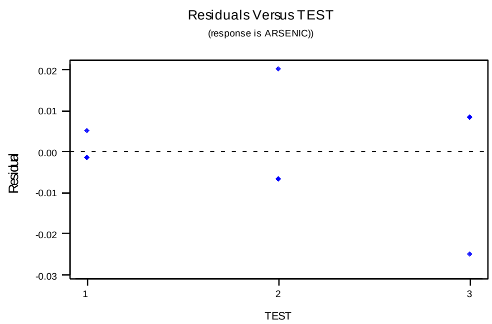
	Subject			
Test	1	2	3	4
Trainee	0.05	0.05	0.04	0.15
Trainer	0.05	0.05	0.04	0.17
Lab	0.04	0.04	0.03	0.10

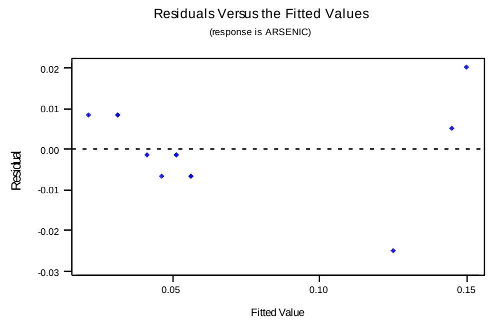
(a)
```
Analysis of Variance for ARSENIC
Source       DF        SS           MS         F       P
TEST          2    0.0014000   0.0007000    3.00   0.125
SUBJECT       3    0.0212250   0.0070750   30.32   0.001
Error         6    0.0014000   0.0002333
Total        11    0.0240250
```

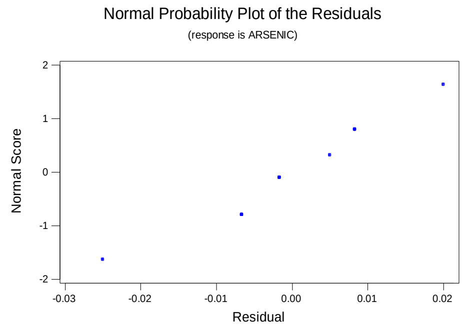
Fail to reject H_0, there is no evidence of differences between the tests.

(b) Some indication of variability increasing with the magnitude of the response.









13-50 An article in the *Food Technology Journal* (1956, Vol. 10, pp. 39–42) described a study on the protopectin content of tomatoes during storage. Four storage times were selected, and samples from nine lots of tomatoes were analyzed. The protopectin content (expressed as hydrochloric acid soluble fraction mg/kg) is in Table 13E-1.

(a) The researchers in this study hypothesized that mean protopectin content would be different at different storage times. Can you confirm this hypothesis with a statistical test using $\alpha = 0.05$?
(b) Find the *P*-value for the test in part (a).
(c) Which specific storage times are different? Would you agree with the statement that protopectin content decreases as storage time increases?
(d) Analyze the residuals from this experiment.

TABLE • 13E-1 Protopectin Content of Tomatoes in Storage

Storage Time	Lot 1	2	3	4	5	6	7	8	9
0 days	1694.0	989.0	917.3	346.1	1260.0	965.6	1123.0	1106.0	1116.0
7 days	1802.0	1074.0	278.8	1375.0	544.0	672.2	818.0	406.8	461.6
14 days	1568.0	646.2	1820.0	1150.0	983.7	395.3	422.3	420.0	409.5
21 days	415.5	845.4	377.6	279.4	447.8	272.1	394.1	356.4	351.2

(a) Analysis of Variance of PROPECTIN

```
Source       DF        SS        MS      F      P
STORAGE       3   1972652    657551   4.33  0.014
LOT           8   1980499    247562   1.63  0.169
Error        24   3647150    151965
Total        35   7600300
```

Reject H_0, and conclude that the storage times affect the mean level of propectin.

(b) P-value = 0.014

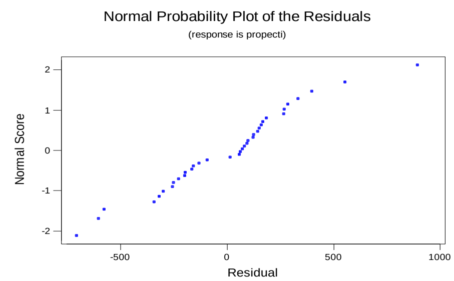
(c)

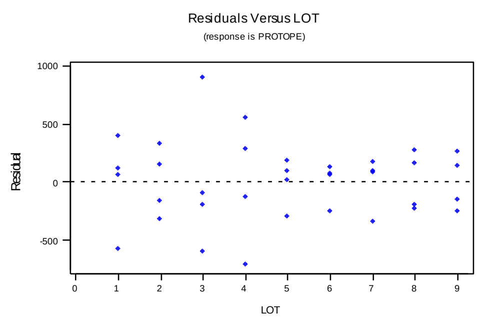
```
Fisher's pairwise comparisons
Family error rate = 0.196
Individual error rate = 0.0500
Critical value = 2.037
Intervals for (column level mean) - (row level mean)
```

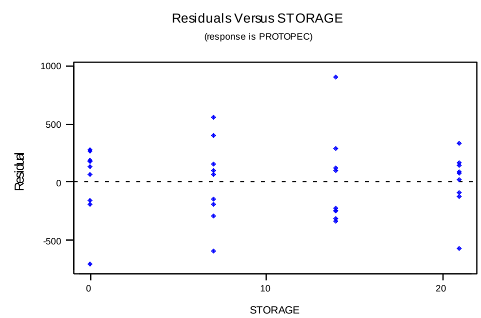
	0	7	14
7	-171		
	634		
14	-214	-445	
	592	360	
21	239	8	50
	1045	813	856

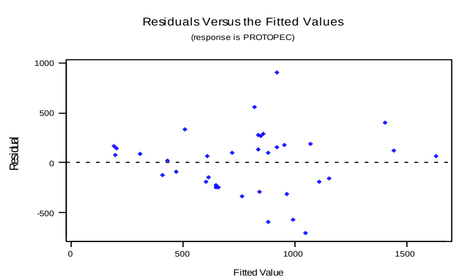
There are differences between 0 and 21 days; 7 and 21 days; and 14 and 21 days. The propectin levels are significantly different at 21 days from the other storage times so there is evidence that the mean level of propectin decreases with storage time. However, differences such as between 0 and 7 days and 7 and 14 days were not significant so that the level is not simply a linear function of storage days.

(d) Observations from lot 3 at 14 days appear unusual. Otherwise, the residuals are acceptable.









13-51 An experiment was conducted to investigate leaking current in a SOS MOSFETS device. The purpose of the experiment was to investigate how leakage current varies as the channel length changes. Four channel lengths were selected. For each channel length, five different widths were also used, and width is to be considered a nuisance factor. The data are as follows:

Channel Length	Width				
	1	2	3	4	5
1	0.7	0.8	0.8	0.9	1.0
2	0.8	0.8	0.9	0.9	1.0
3	0.9	1.0	1.7	2.0	4.0
4	1.0	1.5	2.0	3.0	20.0

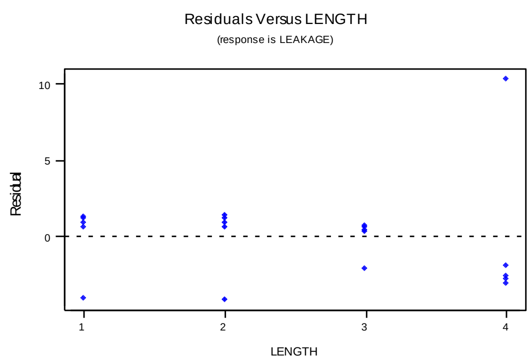
(a) Test the hypothesis that mean leakage voltage does not depend on the channel length using $\alpha = 0.05$.
(b) Analyze the residuals from this experiment. Comment on the residual plots.
(c) The observed leakage voltage for channel length 4 and width 5 was erroneously recorded. The correct observation is 4.0. Analyze the corrected data from this experiment. Is there evidence to conclude that mean leakage voltage increases with channel length?

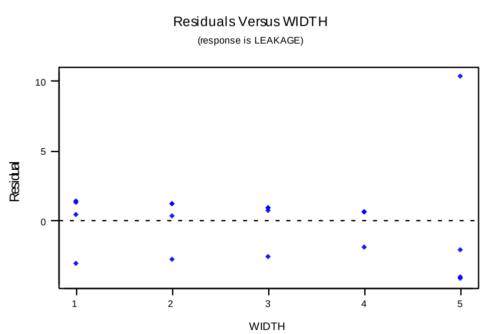
A version of the electronic data file has the reading for length 4 and width 5 as 2. It should be 20.

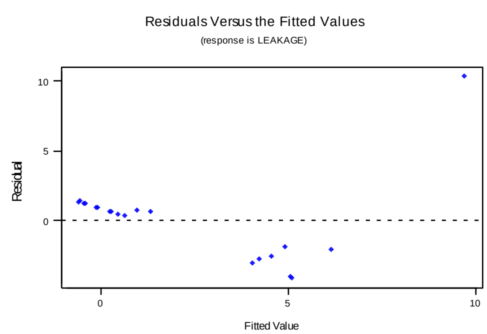
(a)
```
Analysis of Variance for LEAKAGE
Source      DF        SS        MS      F      P
LENGTH       3     72.66     24.22   1.61  0.240
WIDTH        4     90.52     22.63   1.50  0.263
Error       12    180.83     15.07
Total       19    344.01
```

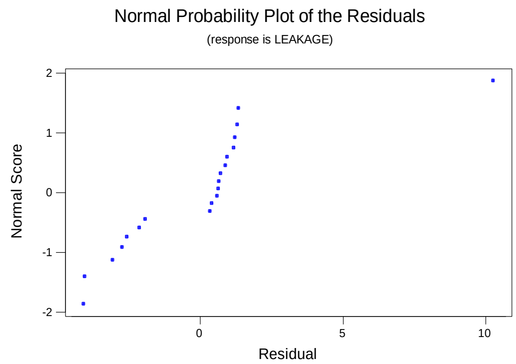
Fail to reject H_0, mean leakage voltage does not depend on the channel length.

(b) One unusual observation in width 5, length 4. There are some problems with the normal probability plot, including the unusual observation.









(c) Analysis of Variance for LEAKAGE VOLTAGE

Source	DF	SS	MS	F	P
LENGTH	3	8.1775	2.7258	6.16	0.009
WIDTH	4	6.8380	1.7095	3.86	0.031
Error	12	5.3100	0.4425		
Total	19	20.3255			

Reject H_0. And conclude that the mean leakage voltage does depend on channel length. By removing the data point that was erroneous, the analysis results in a conclusion. The erroneous data point that was an obvious outlier had a strong effect the results of the experiment.

Supplemental Exercises

13-52 Consider the following computer output.

Source	DF	SS	MS	F	P
Factor	?	?	?	?	?
Error	15	167.5	?		
Total	19	326.2			
S = 3.342	R – Sq = ?	R – Sq(adj) = 34.96%			

(a) How many levels of the factor were used in this experiment?
(b) How many replicates were used?
(c) Fill in the missing information. Use bounds for the *P*-value.
(d) What conclusions would you draw if $\alpha = 0.05$? What if $\alpha = 0.01$?

(a) Note that df(Factor) = df(Total) – df(Error) = 19 – 15 = 4. Because the number of levels for a factor = df(Factor) + 1, 5 levels were used in the experiment.

(b) Total number of observations = df(Total) + 1 = 19 + 1 = 20.
Because there are 5 levels used in this experiment, the number of replicates = 20/5 = 4.

(c) From part (a), the df(Factor) = 4.
SS(Factor) = SS(Total) – SS(Error) = 326.2 – 167.5 = 158.7.
MS(Factor) = SS(Factor)/DF(Factor) = 158.7/4 = 39.675.
MS(Error) = SS(Error)/DF(Error) = 167.5/15 = 11.167.
F = MS(Factor)/MS(Error) = 39.675/11.167 = 3.553
0.025 < P-value < 0.05.

(d) Because the P-value < $\alpha = 0.05$ we reject H_0 for $\alpha = 0.05$. There are significance differences in the factor level means at $\alpha = 0.05$. Because the P-value > $\alpha = 0.01$ we fail to reject H_0 for $\alpha = 0.01$. There are not significance differences in the factor level means at $\alpha = 0.01$.

13-53 Consider the following computer output.

Source	DF	SS	MS	F	P
Factor	?	126.880	63.4401	?	?
Block	?	54.825	18.2751		
Error	6	?	2.7403		
Total	11	198.147			

(a) How many levels of the factor were used in this experiment?
(b) How many blocks were used?
(c) Fill in the missing information. Use bounds for the *P*-value.
(d) What conclusions would you draw if α = 0.05? What if α = 0.01?

(a) Because MS = SS/df(Factor), df(Factor) = SS/MS = 126.880/63.4401 = 2. The number of levels = df(Factor) + 1 = 2 + 1 = 3. Therefore, 3 levels of the factor were used.

(b) Because df(Total) = df(Factor) + df(Block) + df(Error)
11 = 2 + df(Block) + 6. Therefore, df(Block) = 3. Therefore, 4 blocks were used in the experiment.

(c) From parts (a) and (b), df(Factor) = 3 and df(Block) = 2
SS(Error) = df(Error)MS(Error) = (6)2.7403 = 16.4418
F = MS(Factor)/MS(Error) = 63.4401/2.7403 = 23.15
From Appendix Table VI, P-value < 0.01

(d) Because the P-value < 0.01 we reject H_0. There are significant differences in the factor level means at α = 0.05 or α = 0.01.

13-54 An article in *Lubrication Engineering* (December 1990) described the results of an experiment designed to investigate the effects of carbon material properties on the progression of blisters on carbon face seals. The carbon face seals are used extensively in equipment such as air turbine starters. Five different carbon materials were tested, and the surface roughness was measured. The data are as follows:

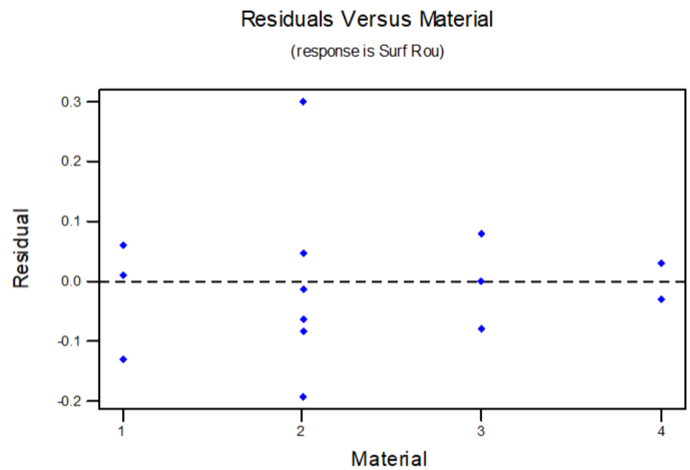
Carbon Material Type	Surface Roughness					
EC10	0.50	0.55	0.55	0.36		
EC10A	0.31	0.07	0.25	0.18	0.56	0.20
EC4	0.20	0.28	0.12			
EC1	0.10	0.16				

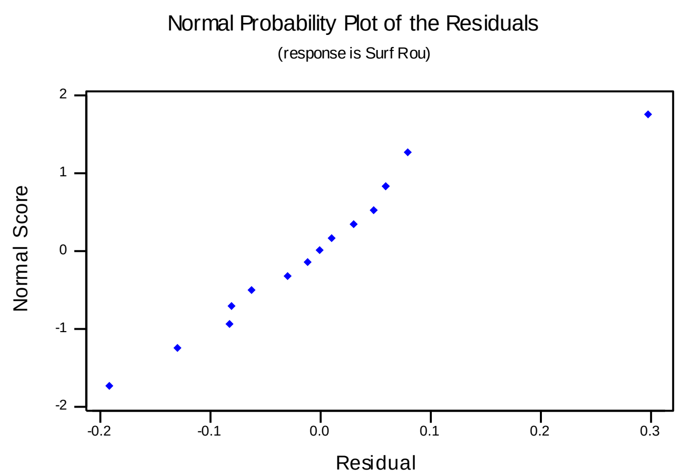
(a) Does carbon material type have an effect on mean surface roughness? Use α = 0.05.
(b) Find the residuals for this experiment. Does a normal probability plot of the residuals indicate any problem with the normality assumption?
(c) Plot the residuals versus $\hat{y}_{ij}$. Comment on the plot.
(d) Find a 95% confidence interval on the difference between mean surface roughness for the EC10 and the EC1 carbon grades.
(e) Apply the Fisher LSD method to this experiment. Summarize your conclusions regarding the effect of material type on surface roughness.

(a)
```
Analysis of Variance for SURFACE ROUGNESS
Analysis of Variance for y
Source      DF        SS        MS        F       P
Material     3    0.2402    0.0801     4.96   0.020
Error       11    0.1775    0.0161
Total       14    0.4177
```

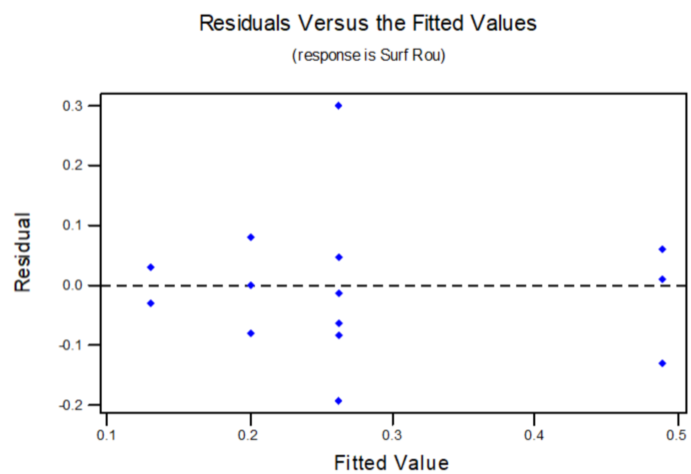
Reject H_0

(b) One observation is an outlier.





(c) There appears to be a problem with constant variance. This may be due to the outlier in the data.



(d) 95% confidence interval on the difference in the means of EC10 and EC1

$$\bar{y}_1 - \bar{y}_4 - t_{0.025,,11}\sqrt{\frac{MS_E}{n_1} + \frac{MS_E}{n_2}} \le \mu_1 - \mu_3 \le \bar{y}_1 - \bar{y}_4 + t_{0.025,,11}\sqrt{\frac{MS_E}{n_1} + \frac{MS_E}{n_2}}$$

$$(0.490 - 0.130) - 2.201\sqrt{\frac{(0.0161)}{4} + \frac{(0.0161)}{2}} \le \mu_1 - \mu_4 \le (0.490 - 0.130) + 2.201\sqrt{\frac{(0.0161)}{4} + \frac{(0.0161)}{2}}$$

$$0.118 \le \mu_1 - \mu_4 \le 0.602$$

13-55 An article in the *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* [(1992, Vol. 15(2), pp. 146–153)] described an experiment in which the contact resistance of a brake-only relay was studied for three different materials (all were silver-based alloys). The data are as follows.

Alloy	Contact Resistance				
1	95	97	99	98	99
	99	99	94	95	98
2	104	102	102	105	99
	102	111	103	100	103
3	119	130	132	136	141
	172	145	150	144	135

(a) Does the type of alloy affect mean contact resistance? Use α = 0.01.
(b) Use Fisher's LSD method to determine which means differ.
(c) Find a 99% confidence interval on the mean contact resistance for alloy 3.
(d) Analyze the residuals for this experiment.

(a)
```
Analysis of Variance for RESISTANCE
Source     DF        SS        MS        F        P
ALLOY       2   10941.8    5470.9    76.09    0.000
Error      27    1941.4      71.9
Total      29   12883.2
```

Reject H_0, the type of alloy has a significant effect on mean contact resistance.

(b)
```
Fisher's pairwise comparisons
 Family error rate = 0.119
 Individual error rate = 0.0500
 Critical value = 2.052
 Intervals for (column level mean) - (row level mean)
                 1            2
      2      -13.58
               1.98
      3      -50.88       -45.08
             -35.32      -29.52
```

There are differences in the mean resistance for alloy types 1 and 3; and types 2 and 3.

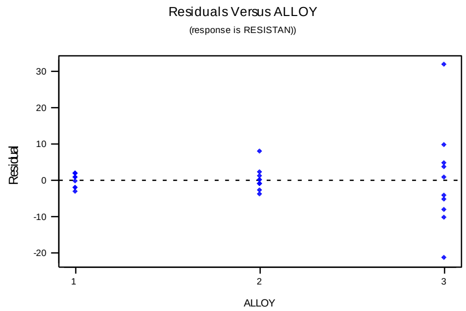
(c) 99% confidence interval on the mean contact resistance for alloy 3

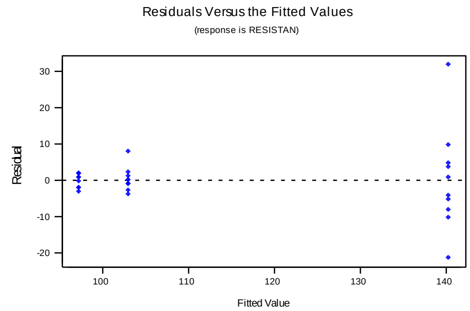
$$\bar{y}_3 - t_{0.005,27}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_3 + t_{0.005,27}\sqrt{\frac{MS_E}{n}}$$

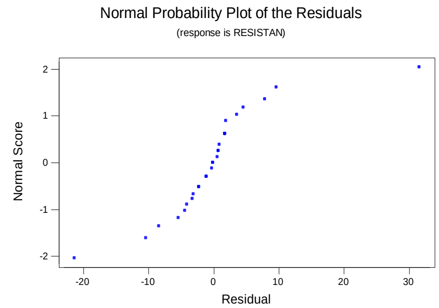
$$140.4 - 2.771\sqrt{\frac{71.9}{10}} \le \mu_3 \le 140.4 + 2.771\sqrt{\frac{71.9}{10}}$$

$$132.97 \le \mu_1 \le 147.83$$

(d) Variability of the residuals increases with the response. The normal probability plot has some curvature in the tails, indicating a problem with the normality assumption. A transformation of the response should be conducted.







13-56 An article in the *Journal of Quality Technology* [(1982, Vol. 14(2), pp. 80–89)] described an experiment in which three different methods of preparing fish were evaluated on the basis of sensory criteria, and a quality score was assigned. Assume that these methods have been randomly selected from a large population of preparation methods. The data are in the following table:

Method	Score			
1	24.4	23.2	25.0	19.7
	22.2	24.4	23.8	18.0
2	22.1	19.5	17.3	19.7
	22.3	23.2	21.4	22.6
3	23.3	22.8	22.4	23.7
	20.4	23.5	20.8	24.1

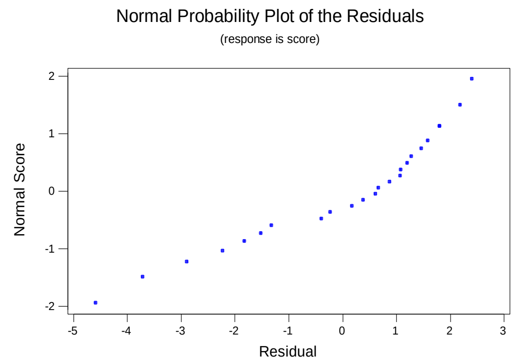
(a) Is there any difference in preparation methods? Use α = 0.05.
(b) Calculate the *P*-value for the *F*-statistic in part (a).
(c) Analyze the residuals from this experiment and comment on model adequacy.
(d) Estimate the components of variance.

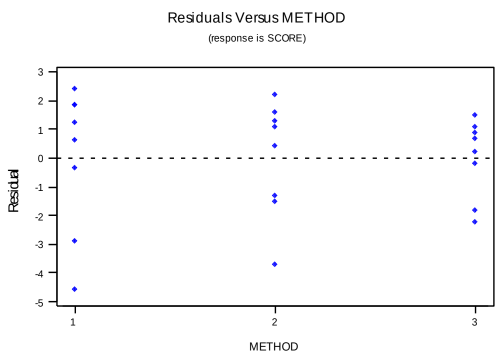
(a)
```
Analysis of Variance for SCORE
Source      DF        SS        MS        F        P
METHOD       2     13.55      6.78     1.68    0.211
Error       21     84.77      4.04
Total       23     98.32
```

Fail to reject H_0

(b) *P*-value = 0.211

(c) There is some curvature in the normal probability plot. There appears to be some differences in the variability for the different methods. The variability for method one is larger than the variability for method 3.





(d) $$\hat{\sigma}_\tau^2 = \frac{MS_{Treatments} - MS_E}{n} = \frac{6.78 - 4.04}{8} = 0.342$$
$$\hat{\sigma}^2 = MS_E = 4.04$$

13-57 An article in the *Journal of Agricultural Engineering Research* (1992, Vol. 52, pp. 53–76) described an experiment to investigate the effect of drying temperature of wheat grain on baking quality bread. Three temperature levels were used, and the response variable measured was the volume of the loaf of bread produced. The data are as follows:

Temperature (°C)	Volume (CC)				
70.0	1245	1235	1285	1245	1235
75.0	1235	1240	1200	1220	1210
80.0	1225	1200	1170	1155	1095

(a) Does drying temperature affect mean bread volume? Use α = 0.01.
(b) Find the *P*-value for this test.
(c) Use the Fisher LSD method to determine which means are different.
(d) Analyze the residuals from this experiment and comment on model adequacy.

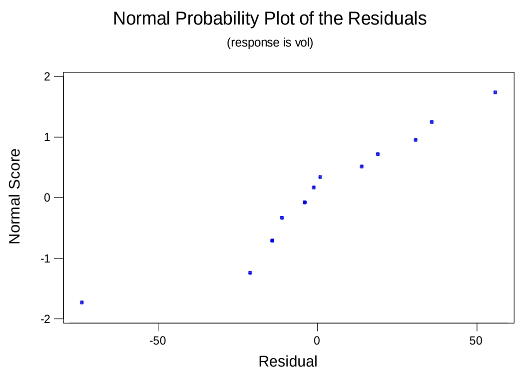
(a)
```
Analysis of Variance for VOLUME
Source       DF       SS       MS      F       P
TEMPERATURE   2    16480     8240   7.84   0.007
Error        12    12610     1051
Total        14    29090
```
Reject H_0.

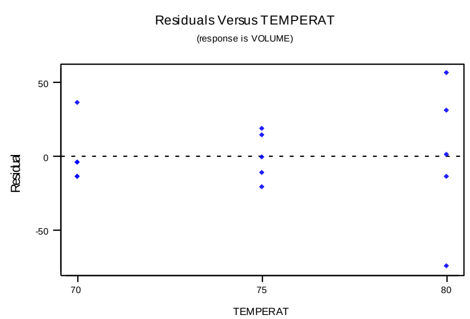
(b) *P-value* = 0.007

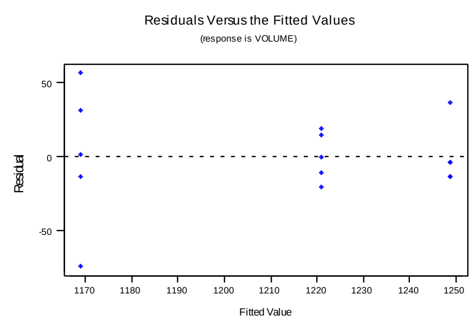
(c)
```
Fisher's pairwise comparisons
  Family error rate = 0.116
  Individual error rate = 0.0500
  Critical value = 2.179
  Intervals for (column level mean) - (row level mean)
                    70            75
       75        -16.7
                  72.7
       80         35.3          7.3
                 124.7         96.7
```

There are significant differences in the mean volume for temperature levels 70 and 80; and 75 and 80. The highest temperature results in the smallest mean volume.

(d) There are some relatively small differences in the variability at the different levels of temperature. The variability decreases with the fitted values. There is an unusual observation on the normal probability plot.







13-58 An article in *Agricultural Engineering* (December 1964, pp. 672–673) described an experiment in which the daily weight gain of swine is evaluated at different levels of housing temperature. The mean weight of each group of swine at the start of the experiment is considered to be a nuisance factor. The data from this experiment are as follows:

Mean Weight (lbs)	Housing Air Temperatures (°F)					
	50	60	70	80	90	100
100	1.37	1.58	2.00	1.97	1.40	0.39
150	1.47	1.75	2.16	1.82	1.14	–0.19
200	1.19	1.91	2.22	1.67	0.88	–0.77

(a) Does housing air temperature affect mean weight gain? Use $\alpha = 0.05$.
(b) Use Fisher's LSD method to determine which temperature levels are different.
(c) Analyze the residuals from this experiment and comment on model adequacy.

(a) Analysis of Variance of Weight Gain

Source	DF	SS	MS	F	P
MEANWEIG	2	0.2227	0.1113	1.48	0.273
AIRTEMP	5	10.1852	2.0370	27.13	0.000
Error	10	0.7509	0.0751		
Total	17	11.1588			

Reject H_0 and conclude that the air temperature has an effect on the mean weight gain.

(b) Fisher's pairwise comparisons

Family error rate = 0.314

Individual error rate = 0.0500

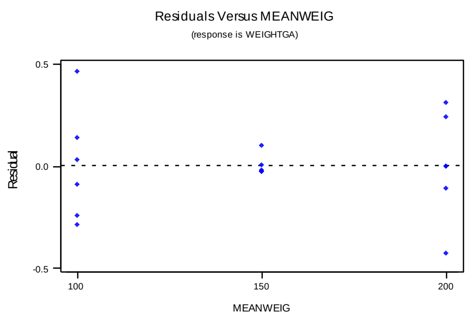
Critical value = 2.179

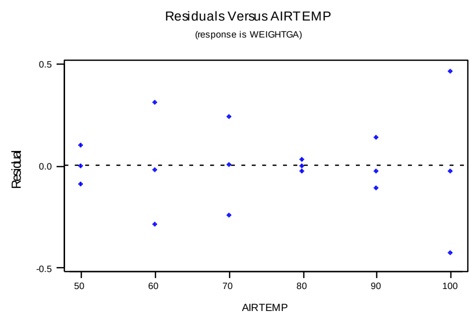
Intervals for (column level mean) - (row level mean)

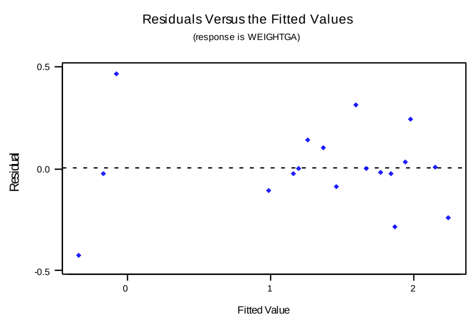
	50	60	70	80	90
60	-0.9101 0.1034				
70	-1.2901 -0.2766	-0.8868 0.1268			
80	-0.9834 0.0301	-0.5801 0.4334	-0.2001 0.8134		
90	-0.3034 0.7101	0.0999 1.1134	0.4799 1.4934	0.1732 1.1868	
100	1.0266 2.0401	1.4299 2.4434	1.8099 2.8234	1.5032 2.5168	0.8232 1.8368

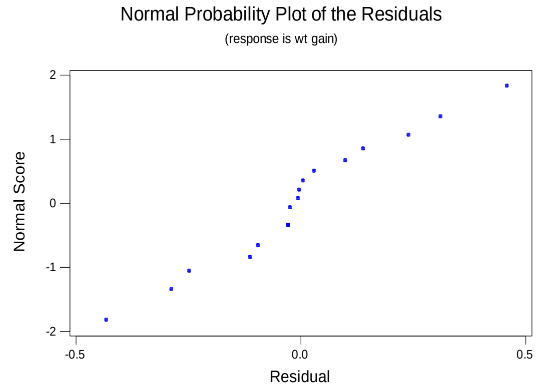
There are significant differences in the mean air temperature levels 50 and 70, 100; 60 and 90, 100; 70 and 90, 100; 80 and 90, 100; and 90 and 100. The mean of temperature level 100 is different from all the other temperatures.

(c) There appears to be some problems with the assumption of constant variance.









13-59 An article in *Communications of the ACM* [(1987, Vol. 30(5), pp. 53–76] reported on a study of different algorithms for estimating software development costs. Six algorithms were applied to eight software development projects and the percent error in estimating the development cost was observed. The data are in Table 13E-2.

(a) Do the algorithms differ in mean cost estimation accuracy? Use α = 0.05.
(b) Analyze the residuals from this experiment.
(c) Which algorithm would you recommend for use in practice?

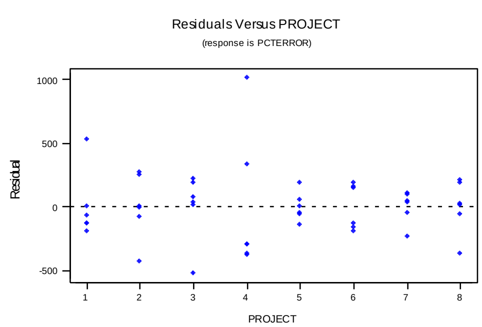
TABLE • 13E-2 Software Development Costs

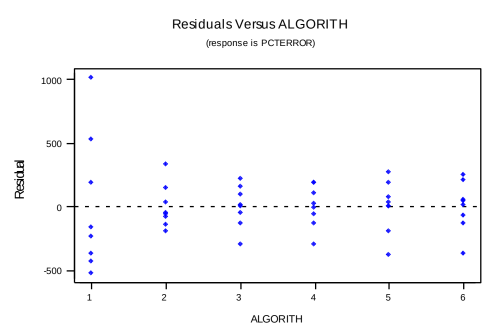
	Project							
Algorithm	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1(SLIM)	1244	21	82	2221	905	839	527	122
2(COCOMO-A)	281	129	396	1306	336	910	473	199
3(COCOMO-R)	220	84	458	543	300	794	488	142
4(COCOMO-C)	225	83	425	552	291	826	509	153
5(FUNCTION POINTS)	19	11	–34	121	15	103	87	–17
6(ESTIMALS)	–20	35	–53	170	104	199	142	41

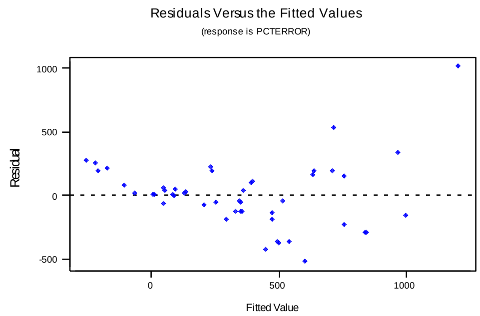
(a)
```
Analysis of Variance for PCTERROR
Source       DF        SS        MS      F      P
ALGORITH      5   2825746    565149   6.23  0.000
PROJECT       7   2710323    387189   4.27  0.002
Error        35   3175290     90723
Total        47   8711358
```

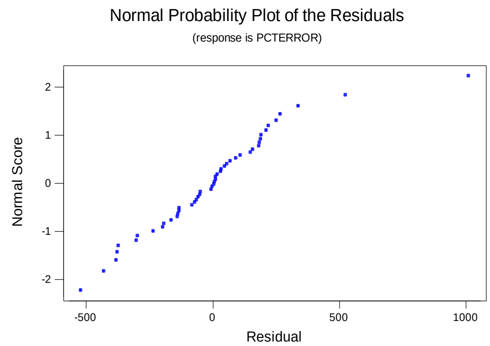
Reject H_0, the algorithms are significantly different.

(b) The residuals look acceptable, except there is one unusual point.









(c) The best choice is algorithm 5 because it has the smallest mean and a low variability.

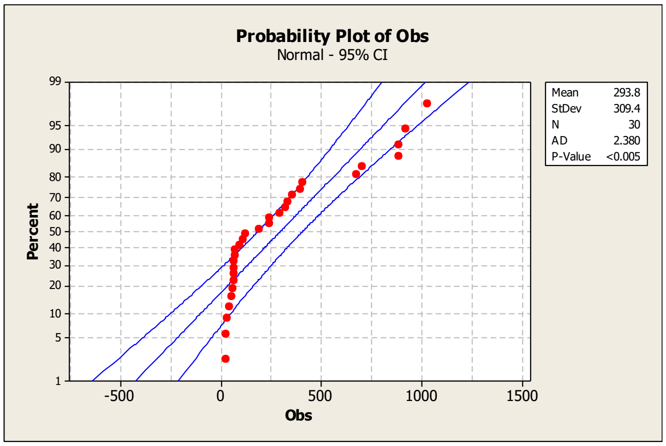
13-60 An article in *Nature Genetics* [(2003, Vol. 34(1), pp. 85–90)], "Treatment-Specific Changes in Gene Expression Discriminate in vivo Drug Response in Human Leukemia Cells," reported the results of a study of gene expression as a function of different treatments for leukemia. Three treatment groups are mercaptopurine (MP) only, low-dose methotrexate (LDMTX) and MP, and high-dose methotrexate (HDMTX) and MP. Each group contained ten subjects. The responses from a specific gene are shown in Table 13E-3.

(a) Check the normality of the data. Can you assume that these samples are from normal populations?
(b) Take the logarithm of the raw data and check the normality of the transformed data. Is there evidence to support the claim that the treatment means differ for the transformed data? Use $\alpha = 0.1$.
(c) Analyze the residuals from the transformed data and comment on model adequacy.

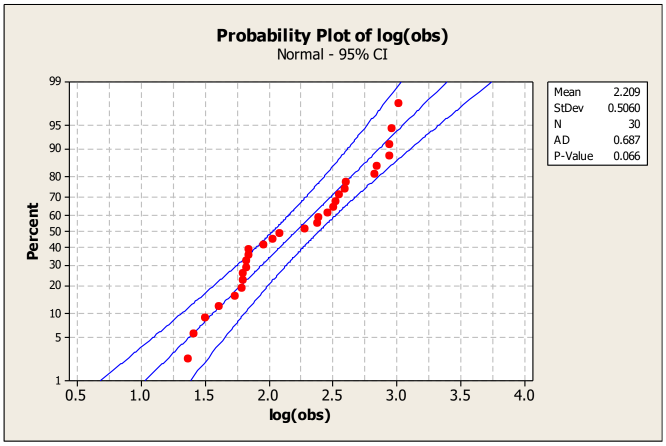
TABLE • 13E-3 Treatment-Specific Changes in Gene Expression

Treatments	Observations									
MP ONLY	334.5	31.6	701	41.2	61.2	69.6	67.5	66.6	120.7	881.9
MP + HDMTX	919.4	404.2	1024.8	54.1	62.8	671.6	882.1	354.2	321.9	91.1
MP + LDMTX	108.4	26.1	240.8	191.1	69.7	242.8	62.7	396.9	23.6	290.4

(a) The normal probability plot shows that the normality assumption is not reasonable.



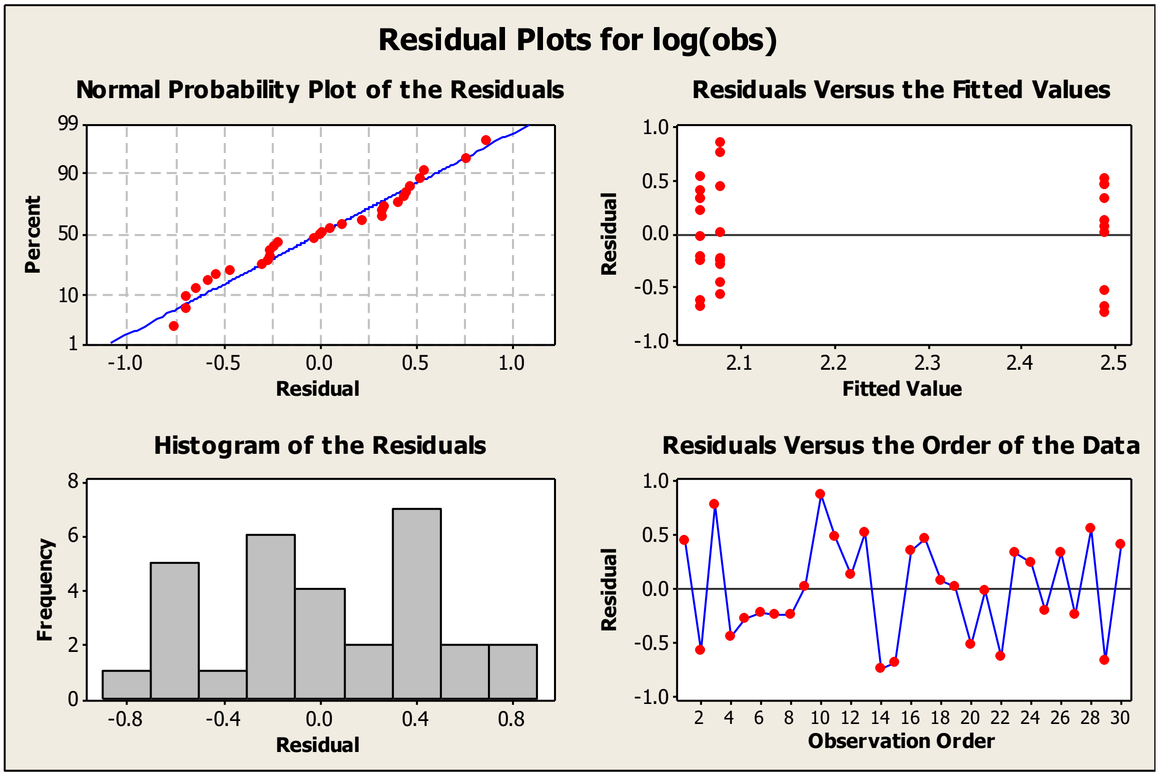
(b) The normal probability plot shows that the normality assumption is reasonable.



Source	DF	SS	MS	F	P
Treatments	2	1.188	0.594	2.57	0.095
Error	27	6.237	0.231		
Total	29	7.425			

There is evidence to support the claim that the treatment means differ at $\alpha = 0.1$ for the transformed data since the P-value = 0.095.

(c) The normal probability plot and the residual plots show that the model assumptions are reasonable.



13-61 Consider an ANOVA situation with $a = 5$ treatments. Let $\sigma^2 = 9$ and $\alpha = 0.05$, and suppose that $n = 4$.

(a) Find the power of the ANOVA *F*-test when $\mu_1 = \mu_2 = \mu_3 = 1$, $\mu_4 = 3$, and $\mu_5 = 2$.
(b) What sample size is required if you want the power of the *F*-test in this situation to be at least 0.90?

(a) μ=1.6, Φ^2 =0.284, Φ =0.5333

Numerator degrees of freedom = $a-1=4=v_1$

Denominator degrees of freedom = $a(n-1)=15=v_2$

From Chart Figure 13-6, $\beta \approx 0.8$ and the power = $1 - \beta = 0.2$

(b)

n	Φ^2	Φ	a(n-1)	β	Power = 1-β
50	3.56	1.89	245	0.05	0.95

The sample size should be approximately n = 50.

13-62 Consider an ANOVA situation with $a = 4$ means $\mu_1 = 1$, $\mu_2 = 5$, $\mu_3 = 8$, and $\mu_5 = 4$. Suppose that $\sigma^2 = 4$, $n = 4$, and $\alpha = 0.05$.

(a) Find the power of the ANOVA *F*-test.
(b) How large would the sample size have to be if you want the power of the *F*-test for detecting this difference in means to be at least 0.90?

(a) μ = (1+5+8+4)/4 = 4.5 and

$$\Phi^2 = \frac{4[(1-4.5)^2 + (5-4.5)^2 + (8-4.5)^2 + (4-4.5)^2]}{4(4)} = 6.25$$

$$\Phi = 2.5$$

Numerator degrees of freedom = $a-1=3=\nu_1$
Denominator degrees of freedom = $a(n-1)=12=\nu_2$
From Figure 13-6, β = 0.05 and the power = 1 - β = 0.95

(b)

n	Φ^2	Φ	a(n-1)	β	Power = 1-β
4	6.25	2.5	12	0.05	0.95
3	4.6875	2.165	8	0.25	0.75

The sample size should be approximately n = 4.

13-63 An article in *Marine Biology* ["Allozymes an Morphometric Characters of Three Species of Mytilus in the Northern and Southern Hemispheres" (1991, Vol. 111, pp. 323–333)] discussed the ratio of the anterior adductor muscle scar length to shell length for shells from five different geographic locations. The following table is part of a much larger data set from their research.

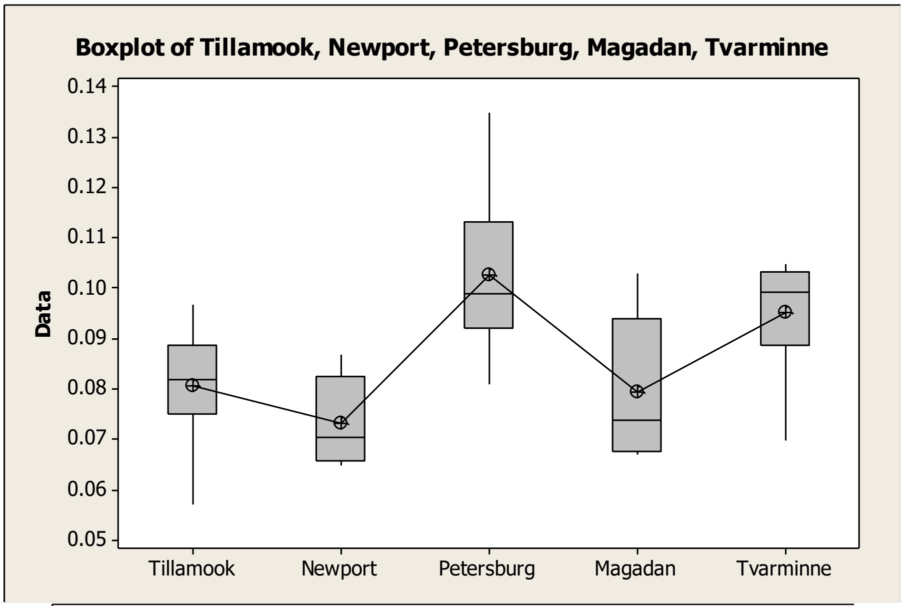
Location	Ratio					
Tillamook, Oregon	0.057	0.081	0.083	0.097	0.081	0.086
Newport, Oregon	0.087	0.066	0.067	0.081	0.074	0.065
Petersburg, Alaska	0.097	0.135	0.081	0.101	0.096	0.106
Magadan, Quebec	0.103	0.091	0.078	0.068	0.067	0.070
Tvarminne, Finland	0.070	0.102	0.095	0.097	0.103	0.105

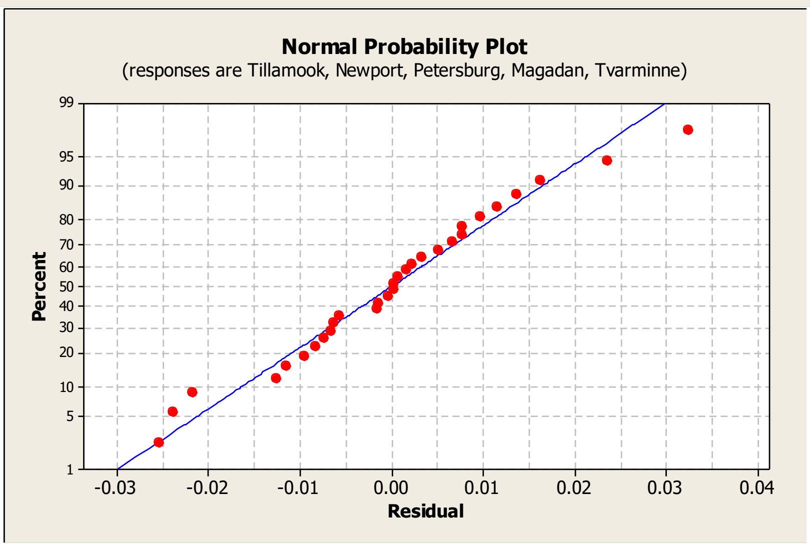
(a) Are there differences in the mean ratios due to different locations at α = 0.05? Calculate the *P*-value.
(b) Analyze the residuals from the experiment. In particular, comment on the normality assumption.

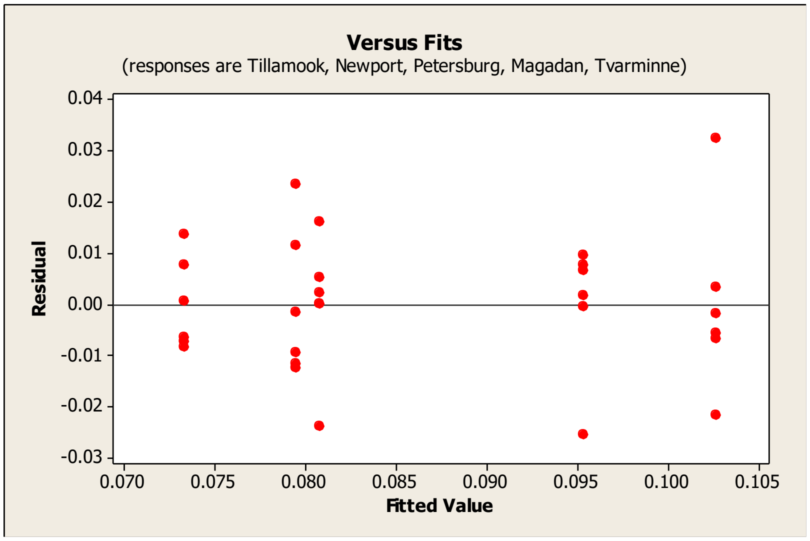
(a)
```
Source   DF        SS        MS     F      P
Factor    4  0.003562  0.000891  4.66  0.006
Error    25  0.004782  0.000191
Total    29  0.008345

S = 0.01383   R-Sq = 42.69%   R-Sq(adj) = 33.52%
Level        N     Mean    StDev
Tillamook    6  0.08083  0.01312
Newport      6  0.07333  0.00905
Petersburg   6  0.10267  0.01792
Magadan      6  0.07950  0.01460
Tvarminne    6  0.09533  0.01297

      Individual 95% CIs For Mean Based on Pooled StDev
Level        ---------+---------+---------+---------+
Tillamook         (-------*-------)
Newport      (-------*-------)
Petersburg                      (------*-------)
Magadan          (-------*-------)
Tvarminne                   (-------*------)
             ---------+---------+---------+---------+
                    0.075     0.090     0.105     0.120

Pooled StDev = 0.01383
```

Because the location P-value = 0.006 < 0.05, the effect of location is significant.

(b) Box plots and residual plots follow. The plots of residuals do not indicate serious departures from assumptions.





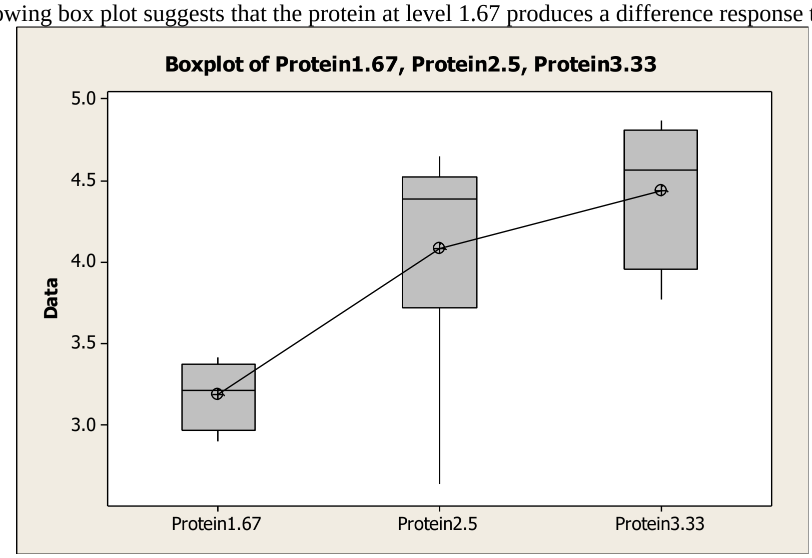


13-64 An article in *Bioresource Technology* ["Preliminary Tests on Nisin and Pediocin Production Using Waste Protein Sources: Factorial and Kinetic Studies" (2006, Vol. 97(4), pp. 605–613)] described an experiment in which pediocin was produced from waste protein. Nisin and pediocin are bacteriocins (compounds produced by bacteria that inhibit related strains) used for food preservation. Three levels of protein (g/L) from trout viscera extracts were compared.

Protein	Pediocin production (ratio to baseline)									
(g/L)	1	2	3	4	5	6	7	8	9	10
1.67	2.90	3.42	3.18	3.25	-	-	-	-	-	-
2.50	2.64	4.52	2.95	4.35	4.65	4.54	4.29	4.42	4.47	3.98
3.33	3.77	4.5	4.62	4.87	-	-	-	-	-	-

(a) Construct box plots of the data. What visual impression do you have from these plots?
(b) Does the level of protein have an effect on mean pediocin production? Use $\alpha = 0.05$.
(c) Would you draw a different conclusion if $\alpha = 0.01$ had been used?
(d) Plot the residuals from the experiment and comment. In particular, comment on the normality of the residuals.
(e) Find a 95% confidence interval on mean pediocin production when the level of protein is 2.50 g/L.

(a) The following box plot suggests that the protein at level 1.67 produces a difference response than the other levels.



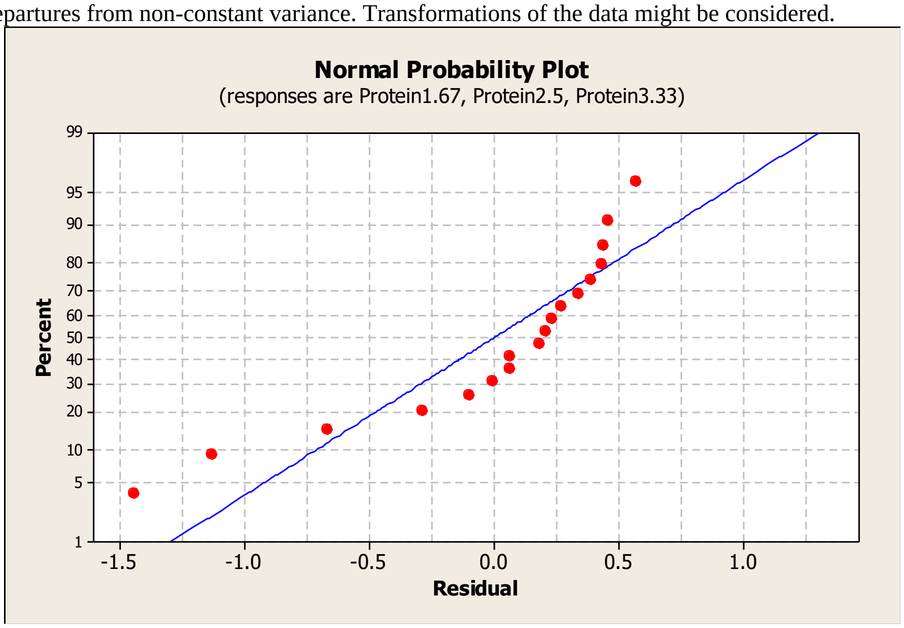
(b)

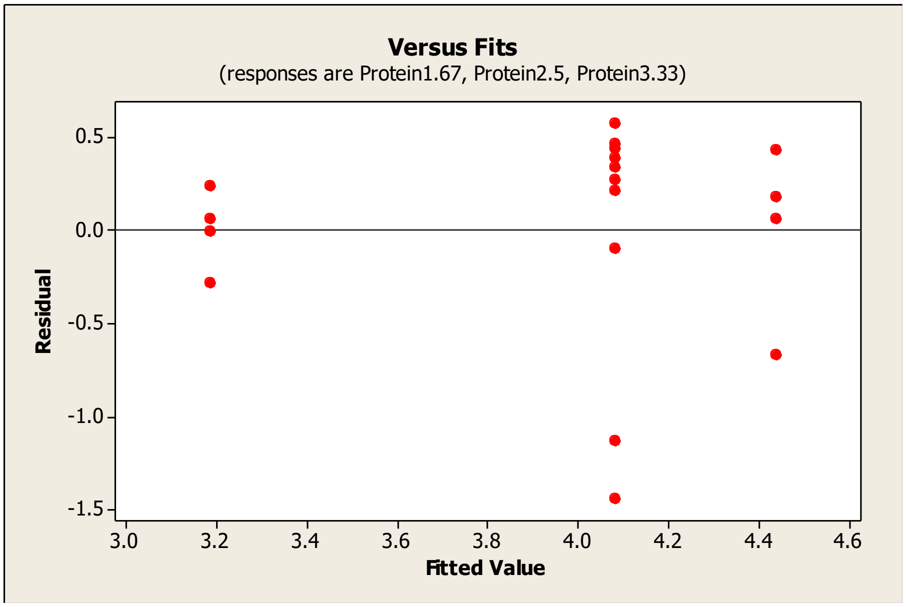
```
Source   DF     SS     MS     F      P
Factor    2  3.455  1.727  4.90  0.023
Error    15  5.286  0.352
Total    17  8.741

S = 0.5936   R-Sq = 39.53%   R-Sq(adj) = 31.46%

                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level         N    Mean   StDev   ----+---------+---------+---------+-----
Protein1.67   4  3.1875  0.2165   (---------*--------)
Protein2.5   10  4.0810  0.7052                    (----*-----)
Protein3.33   4  4.4400  0.4725                     (--------*--------)
                                  ----+---------+---------+---------+-----
                                    2.80      3.50      4.20      4.90

Pooled StDev = 0.5936
```

The P-value for the protein factor = 0.023 < 0.05 so that the level of protein has a significant effect on the pediocin production.

(c) Because the P-value = 0.023 > 0.01 the conclusion would change at $\alpha = 0.01$. At this significance level, the protein level is not significant.

(d) The normality plot has some deviations from a line. Also, the plot of the residuals versus the fitted values indicates some departures from non-constant variance. Transformations of the data might be considered.





(e)

$t_{0.05/2,\,15}$ = 2.1315, and the pooled standard deviation is 0.5936.

$$4.0810 - 2.1315 \times \frac{0.5936}{\sqrt{10}} < \mu_{2.5} < 4.0810 + 2.1315 \times \frac{0.5936}{\sqrt{10}}$$
$$3.681 < \mu_{2.52} < 4.481$$

13-65 Reconsider Exercise 13-9 in which the effect of different coating types on conductivity was investigated. Suppose that the five coating types were selected at random from a large number of types.

(a) How does this change the interpretation of the experiment?
(b) What is an appropriate statistical model for this experiment?
(c) Estimate the parameters of this model.

(a) The experiment now includes random effects.

(b) Random effects model with $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ where τ_i and ε_{ij} are random variables.

(c)

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
1	4	580	145	15.33333
2	4	581	145.25	44.25
3	4	526	131.5	9.666667
4	4	517	129.25	4.25
5	4	581	145.25	7.583333

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1060.5	4	265.125	16.34892	2.41E-05	3.055568
Within Groups	243.25	15	16.21667			
Total	1303.75	19				

$$\hat{\sigma}_{\tau}^{2} = \frac{MS_{Treatments} - MS_E}{n} = \frac{265.125 - 16.21667}{4} = 62.227$$

$$\hat{\sigma}^2 = MS_E = 16.21667$$

13-66 An article in *Journal of Hazardous Materials* ["Toxicity Assessment from Electro-Coagulation Treated-Textile Dye Waste Waters by Bioassays," 2009, Vol. 172(1), pp. 330–337] discussed a study of pollutant removal from textile dyeing waste water with an electro-coagulation technique. Chemical oxygen demand (COD) (a common measure of water pollution) was used as the response, and three different values for electrolysis time were considered. The following data were extracted from a larger study. Suppose that a randomized complete block experiment was conducted with three blocks based on initial pH values.

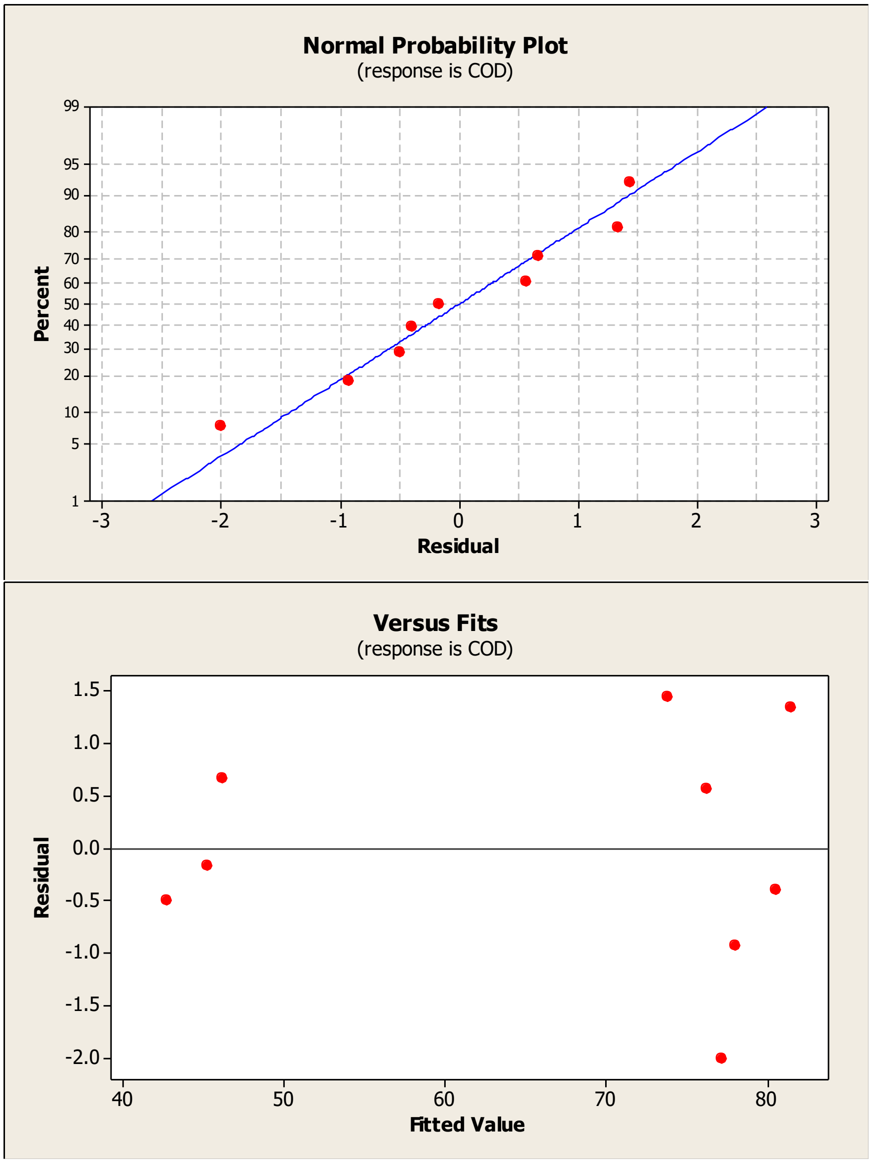
Electrolysis time (min)	**Initial pH**		
	3	**7**	**11**
15	77.1	75.2	42.2
30	80.1	76.8	45.0
45	82.8	75.2	46.8

(a) Is there an effect of electrolysis time at α = 0.05? Calculate the *P*-value.
(b) Analyze the residuals from the experiment.
(c) Calculate a 95% confidence interval on mean COD removal when the electrolysis time is 15 minutes.
(d) Perform an ANOVA assuming that all data are collected at a single pH value. Comment on differences from part (a).

(a)
```
Source  DF       SS       MS       F      P
Time     2    18.81     9.40    3.80  0.119
Block    2  2231.79  1115.89  450.56  0.000
Error    4     9.91     2.48
Total    8  2260.50

S = 1.574   R-Sq = 99.56%   R-Sq(adj) = 99.12%


                Individual 95% CIs For Mean Based on Pooled StDev
Time     Mean     -+---------+---------+---------+--------
15    64.8333     (---------*---------)
30    67.3000               (---------*---------)
45    68.2667                   (---------*---------)
                  -+---------+---------+---------+--------
                62.5      65.0      67.5      70.0
```

Because the P-value for Time = 0.110 > 0.05, there is no significant effect of electrolysis time.

(b)



The residual plots are adequate. Some indication of non-constant variance is present, but there are more points at the greater the fitted values so an increased range of residuals at the greater fitted values is expected.

(c) $t_{0.05/2,4} = 2.7764$, and the pooled standard deviation is 1.574.

$$64.8333 - 2.7764 \times \frac{1.574}{\sqrt{3}} < \mu_{15} < 64.8333 + 2.7764 \times \frac{1.574}{\sqrt{3}}$$
$$62.3103 < \mu_{15} < 67.3564$$

(d)

```
Source  DF    SS   MS     F      P
Time     2    19    9  0.03  0.975
Error    6  2242  374
Total    8  2261
```

```
S = 19.33   R-Sq = 0.83%   R-Sq(adj) = 0.00%


                           Individual 95% CIs For Mean Based on
                           Pooled StDev
Level  N   Mean  StDev  -----+---------+---------+---------+----
15     3  64.83  19.62  (-----------------*-----------------)
30     3  67.30  19.38    (-----------------*-----------------)
45     3  68.27  18.98    (------------------*-----------------)
                        -----+---------+---------+---------+----
                            45        60        75        90

Pooled StDev = 19.33
```

Without blocks, no effect of electrolysis time is detected. The MSError is much greater.

Mind Expanding Exercises

13-67 Show that in the fixed-effects model analysis of variance $E(MS_E) = \sigma^2$. How would your development change if the random-effects model had been specified?

$MS_E = \dfrac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2}{a(n-1)}$ and $y_{ij} = \mu + a_i + \varepsilon_{ij}$. Then $y_{ij} - \bar{y}_i = \varepsilon_{ij} - \bar{\varepsilon}_{i.}$ and

$\dfrac{\sum_{j=1}^{n}(\varepsilon_{ij}-\bar{\varepsilon}_{i.})}{n-1}$ is recognized to be the sample variance of the independent random variables $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in}$.

Therefore, $E = \left[\dfrac{\sum_{j=1}^{n}(\varepsilon_{ij}-\bar{\varepsilon}_{i.})^2}{n-1}\right] = \sigma^2$ and $E(MS_E) = \sum_{i=1}^{a}\dfrac{\sigma^2}{a} = \sigma^2$.

The development would not change if the random effects model had been specified because $y_{ij} - \bar{y}_i = \varepsilon_{ij} - \bar{\varepsilon}_{i.}$ for this model also.

13-68 Consider testing the equality of the means of two normal populations for which the variances are unknown but are assumed to be equal. The appropriate test procedure is the two-sample *t*-test. Show that the two-sample *t*-test is equivalent to the single-factor analysis of variance *F*-test.

The two sample t-test rejects equality of means if the statistic
$t = \dfrac{|\bar{y}_1 - \bar{y}_2|}{s_p\sqrt{\frac{1}{n}+\frac{1}{n}}}$ is too large.

The ANOVA F-test rejects equality of means if $F = \dfrac{n\sum_{i=1}^{2}(\bar{y}_{i.}-\bar{y}_{..})^2}{MS_E}$ is too large.

Now, $F = \dfrac{\frac{n}{2}(\bar{y}_{1.}-\bar{y}_{2.})^2}{MS_E} = \dfrac{(\bar{y}_{1.}-\bar{y}_{2.})^2}{MS_E\frac{2}{n}}$ and $MS_E = s_p^2$.

Consequently, $F = t^2$. Also, the distribution of the square of a t random variable with a(n – 1) degrees of freedom is an F distribution with 1 and $a(n-1)$ degrees of freedom. Therefore, if the critical value for a two-sided *t*-test of size α is t_0, then the tabulated F value for the F test above is t_0^2. Therefore, $t > t_0$ whenever $F = t^2 > t_0$ and the two tests are identical.

13-69 Consider the ANOVA with $a = 2$ treatments. Show that the MS_E in this analysis is equal to the pooled variance estimate used in the two-sample *t*-test.

$MS_E = \dfrac{\sum_{i=1}^{2}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{2(n-1)}$ and $\dfrac{\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{n-1}$ is recognized as the sample standard deviation calculated from the data from population *i*. Then, $MS_E = \dfrac{s_1^2+s_2^2}{2}$ which is the pooled variance estimate used in the t-test.

13-70 Show that the variance of the linear combination

$$\sum_{i=1}^{a} c_i Y_i. \quad \text{is} \quad \sigma^2 \sum_{i=1}^{a} n_i c_i^2$$

$V(\sum_{i=1}^{a} c_i Y_{i.}) = \sum_{i=1}^{a} c_i^2 V(Y_{i.})$ from the independence of $Y_{1.}, Y_{2.}, \ldots, Y_{a.}$.

Also, $V(Y_{i.}) = n_i \sigma_i^2$. Then, $V(\sum_{i=1}^{a} c_i Y_{i.}) = \sigma^2 \sum_{i=1}^{a} c_i^2 n_i$

13-71 In a fixed-effects model, suppose that there are *n* observations for each of four treatments. Let Q_1^2, Q_2^2, and Q_3^2 be single-degree-of-freedom sums of squares for orthogonal contrasts. A contrast is a linear combination of the treatment means with coefficients that sum to zero. The coefficient vectors of orthogonal contrasts are orthogonal vectors. Prove that $SS_{\text{Treatments}} = Q_1^2 + Q_2^2 + Q_3^2$.

If b, c, and d are the coefficients of three orthogonal contrasts, it can be shown that

$\dfrac{(\sum_{i=1}^{4} b_i y_{i.})^2}{\sum_{i=1}^{a} b_i^2} + \dfrac{(\sum_{i=1}^{a} c_i y_{i.})^2}{\sum_{i=1}^{a} c_i^2} + \dfrac{(\sum_{i=1}^{a} d_i y_{i.})^2}{\sum_{i=1}^{a} d_i^2} = \sum_{i=1}^{a} y_{i.}^2 - \dfrac{(\sum_{i=1}^{a} y_{i.})^2}{a}$ always holds. Upon dividing both sides by n, we have

$Q_1^2 + Q_2^2 + Q_3^2 = \sum_{i=1}^{a} \dfrac{y_i^2}{n} - \dfrac{y_{..}^2}{N}$ which equals $SS_{\text{Treatments}}$.

The equation above can be obtained from a geometrical argument. The square of the distance of any point in four-dimensional space from the zero point can be expressed as the sum of the squared distance along four orthogonal axes. Let one of the axes be the 45 degree line and let the point be ($y_{1.}, y_{2.}, y_{3.}, y_{4.}$). The three orthogonal contrasts are the other three axes. The square of the distance of the point from the origin is $\sum_{i=1}^{a} y_{i.}^2$ and this equals the sum of the squared distances along each of the four axes.

13-72 Consider the single-factor completely randomized design with *a* treatments and *n* replicates. Show that if the difference between any two treatment means is as large as D, the minimum value that the OC curve parameter Φ^2 can take is

$$\Phi^2 = \frac{nD^2}{2a\sigma^2}$$

Because $\Phi^2 = \dfrac{n\sum_{i=1}^{a}(\mu_i - \bar{\mu})^2}{a\sigma^2}$, we only need to shows that $\dfrac{D^2}{2} \le \sum_{i=1}^{a}(\mu_i - \bar{\mu})^2$.

Let μ_1 and μ_2 denote the means that differ by D. Now, $(\mu_1 - x)^2 + (\mu_2 - x)^2$ is minimized for x equal to the mean of μ_1 and μ_2. Therefore, $(\mu_1 - \frac{\mu_1+\mu_2}{2})^2 + (\mu_2 - \frac{\mu_1+\mu_2}{2})^2 \le (\mu_1 - \overline{\mu})^2 + (\mu_2 - \overline{\mu})^2 \le \sum_{i=1}^{a} (\mu_i - \overline{\mu})^2$

Then, $\left(\frac{\mu_1-\mu_2}{2}\right)^2 + \left(\frac{\mu_2-\mu_1}{2}\right)^2 = \frac{D^2}{4} + \frac{D^2}{4} = \frac{D^2}{2} \le \sum_{i=1}^{a} (\mu_i - \overline{\mu})^2$.

13-73 Consider the single-factor completely randomized design. Show that a $100(1-\alpha)$ percent confidence interval for σ^2 is

$$\frac{(N-a)MS_E}{\chi^2_{\alpha/2,N-a}} \le \sigma^2 \le \frac{(N-a)MS_E}{\chi^2_{1-\alpha/2,N-a}}$$

where N is the total number of observations in the experimental design.

$$MS_E = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{a(n-1)} = \frac{\sum_{i=1}^{a} s_i^2}{a} \text{ where } s_i^2 = \frac{\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{n-1}.$$

Because s_i^2 is the sample variance of $y_{i1}, y_{i2}, \ldots, y_{in}$, $\frac{(n-1)S_i^2}{\sigma^2}$ has a chi-square distribution with n-1 degrees of freedom. Then, $\frac{a(n-1)MS_E}{\sigma^2}$ is a sum of independent chi-square random variables. Consequently, $\frac{a(n-1)MS_E}{\sigma^2}$ has a chi-square distribution with a(n – 1) degrees of freedom. Consequently,

$$P(\chi^2_{1-\frac{\alpha}{2},a(n-1)} \le \frac{a(n-1)MS_E}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},a(n-1)}) = 1-\alpha$$

$$= P\left(\frac{a(n-1)MS_E}{\chi^2_{\frac{\alpha}{2},a(n-1)}} \le \sigma^2 \le \frac{a(n-1)MS_E}{\chi^2_{1-\frac{\alpha}{2},a(n-1)}}\right)$$

Using the fact that a(n – 1) = N – a completes the derivation.

13-74 Consider the random-effects model for the singlefactor completely randomized design. Show that a $100(1-\alpha)\%$ confidence interval on the ratio of variance components σ_τ^2/σ^2 is given by

$$L \le \frac{\sigma_\tau^2}{\sigma^2} \le U$$

where

$$L = \frac{1}{n}\left[\frac{MS_{\text{Treatments}}}{MS_E}\left(\frac{1}{f_{\alpha/2,a-1,N-a}}\right) - 1\right]$$

and

$$U = \frac{1}{n}\left[\frac{MS_{\text{Treatments}}}{MS_E}\left(\frac{1}{f_{1-\alpha/2,a-1,N-a}}\right) - 1\right]$$

From the previous exercise, $\frac{(N-a)MS_E}{\sigma^2}$ has a chi-square distribution with $N - a$ degrees of freedom. Now,

$V(\overline{Y}_{i.}) = \sigma_\tau^2 + \frac{\sigma^2}{n}$ and mean square treatment = MS_T is n times the sample variance of $\bar{y}_{1.}, \bar{y}_{2.}, \ldots, \bar{y}_{a.}$.

Therefore, $\frac{(a-1)MS_T}{n(\sigma_\tau^2 + \frac{\sigma^2}{n})} = \frac{(a-1)MS_T}{n\sigma_\tau^2 + \sigma^2}$ has a chi-squared distribution with $a - 1$ degrees of freedom. Using the independence of MS_T and MS_E, we conclude that $\left(\frac{MS_T}{n\sigma_\tau^2 + \sigma^2}\right) / \left(\frac{MS_E}{\sigma^2}\right)$ has an $F_{(a-1),(N-a)}$ distribution.

Therefore,

$$P(f_{1-\frac{\alpha}{2},a-1,N-a} \le \frac{MS_T}{MS_E}\frac{\sigma^2}{n\sigma_\tau^2 + \sigma^2} \le f_{\frac{\alpha}{2},a-1,N-a}) = 1-\alpha$$

$$= P\left(\frac{1}{n}\left[\frac{1}{f_{\frac{\alpha}{2},a-1,N-a}}\frac{MS_T}{MS_E} - 1\right] \le \frac{\sigma_\tau^2}{\sigma^2} \le \frac{1}{n}\left[\frac{1}{f_{1-\frac{\alpha}{2},a-1,N-a}}\frac{MS_T}{MS_E} - 1\right]\right)$$

by an algebraic solution for $\frac{\sigma_\tau^2}{\sigma^2}$ and $P(L \le \frac{\sigma_\tau^2}{\sigma^2} \le U)$.

13-75 Consider a random-effects model for the single-factor completely randomized design.

(a) Show that a 100(1 − α)% confidence interval on the ratio $\sigma_\tau^2 / \sigma^2 + \sigma_\tau^2$ is

$$\frac{L}{L+1} \le \frac{\sigma_\tau^2}{\sigma^2 + \sigma_\tau^2} \le \frac{U}{U+1}$$

where L and U are as defined in Exercise 13-74.

(b) Use the results of part (a) to find a 100(1 − α)% confidence interval for $\sigma^2 / \sigma^2 + \sigma_\tau^2$.

(a) As in the previous exercise, $\frac{MS_T}{MS_E}\frac{\sigma^2}{n\sigma_\tau^2 + \sigma^2}$ has an $F_{(a-1),(N-a)}$ distribution.

and

$$1-\alpha = P(L \le \frac{\sigma_\tau^2}{\sigma^2} \le U)$$

$$= P(\frac{1}{U} \le \frac{\sigma^2}{\sigma_\tau^2} \le \frac{1}{L})$$

$$= P(\frac{1}{U} + 1 \le \frac{\sigma^2}{\sigma_\tau^2} + 1 \le \frac{1}{L} + 1)$$

$$= P(\frac{L}{L+1} \le \frac{\sigma_\tau^2}{\sigma^2 + \sigma_\tau^2} \le \frac{U}{U+1})$$

(b)

$$1-\alpha = P\left(L \le \frac{\sigma_\tau^2}{\sigma^2} \le U\right)$$
$$= P\left(L+1 \le \frac{\sigma_\tau^2+1}{\sigma^2} \le U+1\right)$$
$$= P\left(L+1 \le \frac{\sigma_\tau^2+\sigma^2}{\sigma^2} \le U+1\right)$$
$$= P\left(\frac{1}{U+1} \le \frac{\sigma^2}{\sigma^2+\sigma_\tau^2} \le \frac{1}{L+1}\right)$$

Therefore, $\left(\frac{1}{U+1}, \frac{1}{L+1}\right)$ is a confidence interval for $\frac{\sigma^2}{\sigma_\tau^2+\sigma^2}$

13-76 Consider the fixed-effects model of the completely randomized single-factor design. The model parameters are restricted by the constraint $\sum_{i=1}^{a}\tau_i = 0.$ (Actually, other restrictions could be used, but this one is simple and results in intuitively pleasing estimates for the model parameters.) For the case of unequal sample size $n_1, n_2, \ldots, n_a$, the restriction is $\sum_{i=1}^{a} n_i\tau_i = 0.$ Use this to show that

$$E(MS_{\text{Treatments}}) = \sigma^2 + \frac{\sum_{i=1}^{a} n_i\tau_i^2}{a-1}$$

Does this suggest that the null hypothesis in this model is $H_0: n_1\tau_1 = n_1\tau_2 = \ldots = n_a\tau_a = 0$?

$$MS_T = \frac{\sum_{i=1}^{a} n_i(\bar{y}_{i.}-\bar{y}_{..})^2}{a-1}$$ and for any random variable X, $E(X^2) = V(X) + [E(X)]^2$.

Then,

$$E(MS_T) = \frac{\sum_{i=1}^{a} n_i\{V(\bar{Y}_{i.}-\bar{Y}_{..}) + [E(\bar{Y}_{i.}-\bar{Y}_{..})]^2\}}{a-1}$$

Now, $\bar{Y}_{1.}-\bar{Y}_{..} = (\frac{1}{n_1}-\frac{1}{N})Y_{11} + \ldots + (\frac{1}{n_1}-\frac{1}{N})Y_{1n_1} - \frac{1}{N}Y_{21} - \ldots - \frac{1}{N}Y_{2n_2} - \ldots - \frac{1}{N}Y_{a1} - \ldots - \frac{1}{N}Y_{an_a}$

and

$$V(\bar{Y}_{1.}-\bar{Y}_{..}) = \left((\tfrac{1}{n_1}-\tfrac{1}{N})^2 n_1 + \frac{N-n_1}{N^2}\right)\sigma^2 = (\tfrac{1}{n_1}-\tfrac{1}{N})\sigma^2$$

$$E(\bar{Y}_{1.}-\bar{Y}_{..}) = (\tfrac{1}{n_1}-\tfrac{1}{N})n_1\tau_1 - \frac{n_2}{N}\tau_2 - \ldots - \frac{n_a}{N}\tau_a = \tau_1 \text{from the constraint}$$

Then,

$$E(MS_T) = \frac{\sum_{i=1}^{a} n_i\{(\frac{1}{n_i} - \frac{1}{N})\sigma^2 + \tau_i^2\}}{a-1} = \frac{\sum_{i=1}^{a} [(1 - \frac{n_i}{N})\sigma^2 + n_i\tau_i^2]}{a-1}$$

$$= \sigma^2 + \frac{\sum_{i=1}^{a} n_i\tau_i^2}{a-1}$$

Because $E(MS_E) = \sigma^2$, this does suggest that the null hypothesis is as given in the exercise.

13-77 **Sample Size Determination**. In the single-factor completely randomized design, the accuracy of a100(1 − α)% confidence interval on the difference in any two treatment means is $t_{\alpha/2,a(n-1)}\sqrt{2MS_E/n}$.

(a) Show that if A is the desired accuracy of the interval, the sample size required is

$$n = \frac{2F_{\alpha/2,1,a(n-1)}MS_E}{A^2}$$

(b) Suppose that in comparing $a = 5$ means you have a preliminary estimate of σ^2 of 4. If you want the 95% confidence interval on the difference in means to have an accuracy of 2, how many replicates should you use?

(a) If A is the accuracy of the interval, then $t_{\frac{\alpha}{2},a(n-1)}\sqrt{\frac{2MS_E}{n}} = A$

Squaring both sides yields $t^2_{\frac{\alpha}{2},a(n-1)} \frac{2MS_E}{n} = A^2$

Also, $t^2_{\frac{\alpha}{2},a(n-1)} = F_{\alpha,1,a(n-1)}$. Then,

$$n = \frac{2MS_E F_{\alpha,1,a(n-1)}}{A^2}$$

(b) Because n determines one of the degrees of freedom of the tabulated F value on the right-side of the equation in part (a), some approximation is needed. Because the value for a 95% confidence interval based on a normal distribution is 1.96, we approximate $t_{\frac{\alpha}{2},a(n-1)}$ by 2 and we approximate

$t^2_{\frac{\alpha}{2},a(n-1)} = F_{\alpha,1,a(n-1)}$ by 4.

Then, $n = \frac{2(4)(4)}{4} = 8$. With $n = 8$, $a(n - 1) = 35$ and $F_{0.05,1,35} = 4.12$.

The value 4.12 can be used for F in the equation for n and a new value can be computed for n as

$$n = \frac{2(4)(4.12)}{4} = 8.24 \cong 8$$

Because the solution for n did not change, we can use $n = 8$. If needed, another iteration could be used to refine the value of n.

CHAPTER 14

Section 14-3

14-1 An article in *Industrial Quality Control* (1956, pp. 5–8) describes an experiment to investigate the effect of two factors (glass type and phosphor type) on the brightness of a television tube. The response variable measured is the current (in microamps) necessary to obtain a specified brightness level. The data are shown in the following table:

(a) State the hypotheses of interest in this experiment.
(b) Test the hypotheses in part (a) and draw conclusions using the analysis of variance with α = 0.05.
(c) Analyze the residuals from this experiment.

Glass Type	Phosphor Type		
	1	2	3
1	280	300	290
	290	310	285
	285	295	290
2	230	260	220
	235	240	225
	240	235	230

(a) 1. $H_0: \tau_1 = \tau_2 = 0$ H_1: at least one $\tau_i \neq 0$

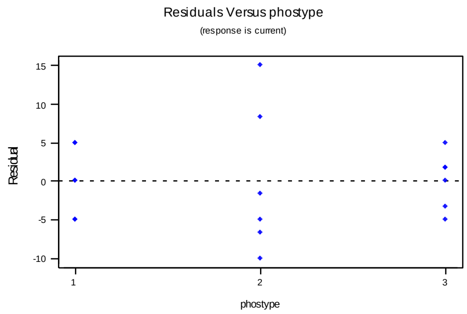
2. $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ H_1: at least one $\beta_j \neq 0$

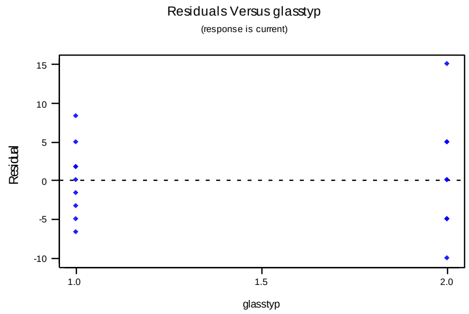
3. $H_0: \tau\beta_{11} = \tau\beta_{12} = ... = \tau\beta_{23} = 0$ H_1: at least one $\tau\beta_{ij} \neq 0$

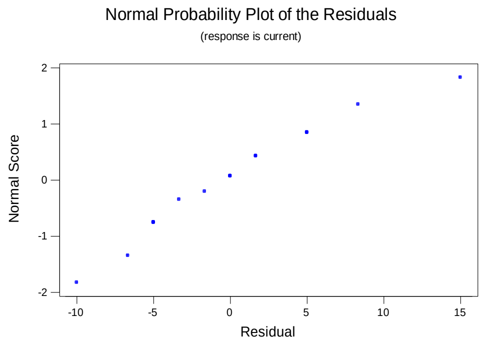
(b)
```
Analysis of Variance for current
Source              DF       SS        MS        F      P
glasstyp             1    14450.0   14450.0   273.79  0.000
phostype             2      933.3     466.7     8.84  0.004
glasstyp*phostype    2      133.3      66.7     1.26  0.318
Error               12      633.3      52.8
Total               17    16150.0
```

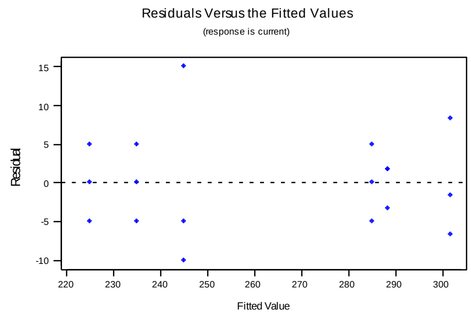
Main effects are significant, but the interaction is not significant. Glass type 1 and phosphor type 2 lead to the highest mean current (brightness).

(c) There appears to be more slightly variability at phosphor type 2 and glass type 2. The normal plot of the residuals indicates that the assumption of normality is reasonable.









14-2 An engineer suspects that the surface finish of metal parts is influenced by the type of paint used and the drying time. He selected three drying times—20, 25, and 30 minutes—and used two types of paint. Three parts are tested with each combination of paint type and drying time. The data are as follows:

	Drying Time (min)		
Paint	**20**	**25**	**30**
1	74	73	78
	64	61	85
	50	44	92
2	92	98	66
	86	73	45
	68	88	85

(a) State the hypotheses of interest in this experiment.
(b) Test the hypotheses in part (a) and draw conclusions using the analysis of variance with α = 0.05.
(c) Analyze the residuals from this experiment.

(a) $H_0 : \tau_1 = \tau_2 = 0$

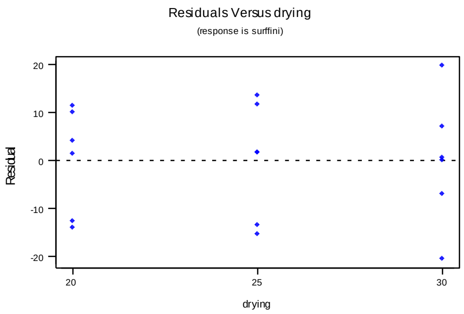
$H_1 : \text{at least one } \tau_j \neq 0$

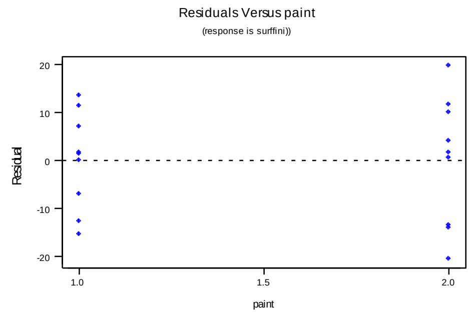
(b) Analysis of variance for SURFACE FINISH

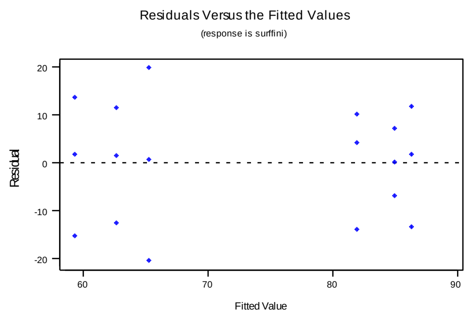
```
Source          DF       SS        MS       F       P
paint            1     355.6     355.6    1.90   0.193
drying           2      27.4      13.7    0.07   0.930
paint*drying     2    1878.8     939.4    5.03   0.026
Error           12    2242.7     186.9
Total           17    4504.4
```

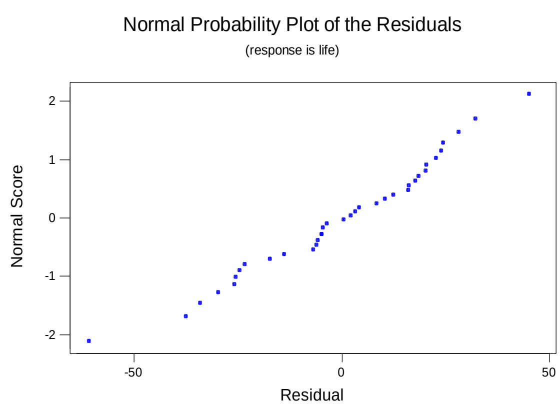
Only the interaction between the paint and drying time is significant.

(c) The residual plots appear reasonable.









14-3 In the book *Design and Analysis of Experiments*,8th edition (2012, John Wiley & Sons), the results of an experiment involving a storage battery used in the launching mechanism of a shoulder-fired ground-to-air missile were presented. Three material types can be used to make the battery plates. The objective is to design a battery that is relatively unaffected by the ambient temperature. The output response from the battery is effective life in hours. Three temperature levels are selected, and a factorial experiment with four replicates is run. The data are as follows:

	Temperature (°F)					
Material	**Low**		**Medium**		**High**	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

(a) Test the appropriate hypotheses and draw conclusions using the analysis of variance with α = 0.05.
(b) Graphically analyze the interaction.
(c) Analyze the residuals from this experiment.

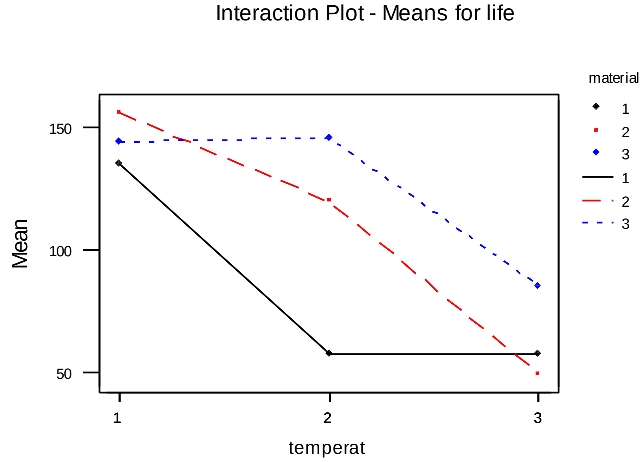
(a) $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ $\quad H_1$: at least one $\tau_j \neq 0$

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ $\quad H_1$: at least one $\beta_j \neq 0$

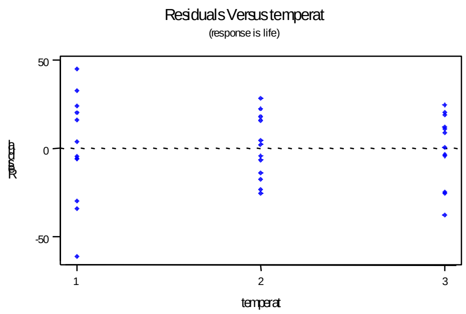
```
Analysis of Variance for life
  Source                  DF        SS        MS       F      P
  material                 2   10683.7    5341.9    7.91  0.002
  temperat                 2   39118.7   19559.4   28.97  0.000
  material*temperat        4    9613.8    2403.4    3.56  0.019
  Error                   27   18230.7     675.2
  Total                   35   77647.0
```

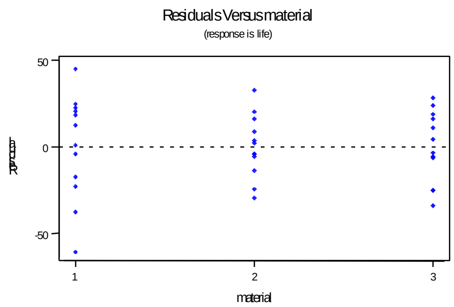
Main factors and the interaction are all significant effects.

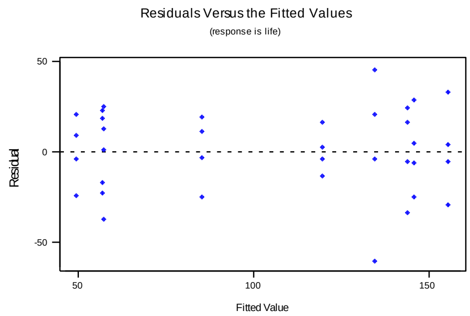
(b) The mean life for material 2 is the highest at temperature level 1, in the middle at temperature level 2 and the lowest at temperature level 3. This shows that there is an interaction.

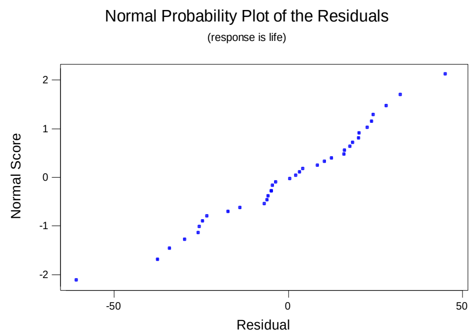


(c) There appears to be slightly more variability at temperature 1 and material 1. The normal probability plot shows that the assumption of normality is reasonable.









14-4 An experiment was conducted to determine whether either firing temperature or furnace position affects the baked density of a carbon anode. The data are as follows:

	Temperature (°C)		
Position	**800**	**825**	**850**
1	570	1063	565
	565	1080	510
	583	1043	590
2	528	988	526
	547	1026	538
	521	1004	532

(a) State the hypotheses of interest.
(b) Test the hypotheses in part (a) using the analysis of variance with α = 0.05. What are your conclusions?
(c) Analyze the residuals from this experiment.
(d) Using Fisher's LSD method, investigate the differences between the mean baked anode density at the three different levels of temperature. Use α = 0.05.

(a) 1. $H_0 : \tau_1 = \tau_2 = 0$

$H_1 :$ at least one $\tau_j \neq 0$

2. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

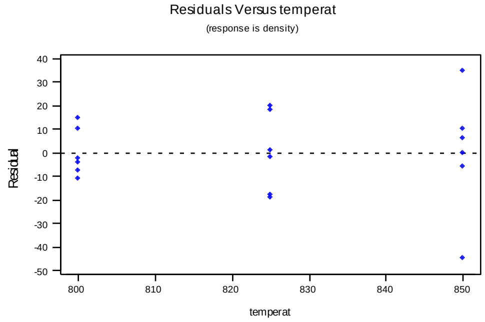
$H_1 :$ at least one $\beta_j \neq 0$

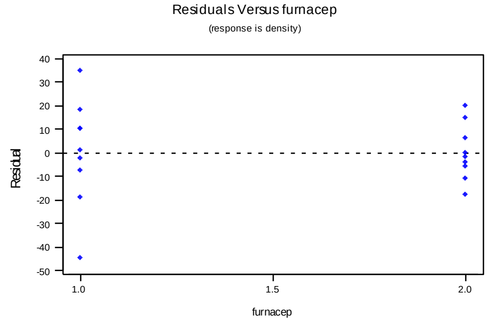
(b)Analysis of Variance for DENSITY

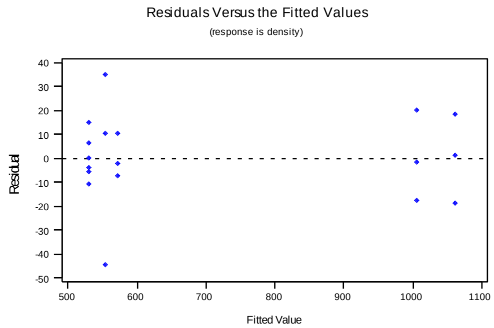
```
Source              DF        SS        MS        F       P
furnacep             1      7160      7160    16.00   0.002
temperat             2    945342    472671  1056.12   0.000
furnacep*temperat    2       818       409     0.91   0.427
Error               12      5371       448
Total               17    958691
```

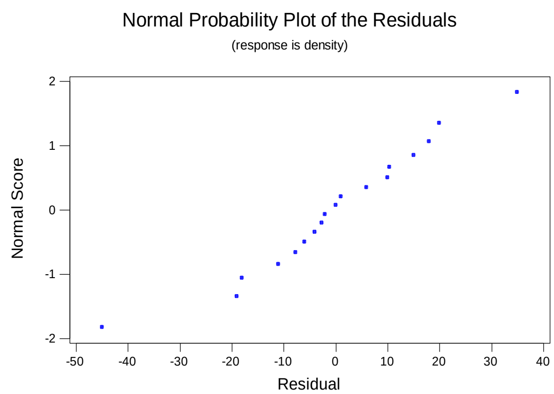
Reject H_0 for both main effects and conclude that both factors are significant.

(c) There appears to be more variability at position 1 and at the highest temperature level. There are two unusual points in the data.









(d)

```
Fisher's pairwise comparisons
Family error rate = 0.117
Individual error rate = 0.0500
Critical value = 2.131
Intervals for (column level mean) - (row level mean)
               800          825
 825        -518.4
            -445.0
 850         -27.9        453.8
              45.5        527.2
```

There are significant differences in the temperature levels 800 and 825, and 825 and 850. Therefore, temperature level 825 is different from the other two levels.

14-5 An article in *Technometrics* ["Exact Analysis of Means with Unequal Variances" (2002, Vol. 44, pp. 152–160)] described the technique of the analysis of means (ANOM) and presented the results of an experiment on insulation. Four insulation types were tested at three different temperatures. The data are as follows:

(a) Write a model for this experiment.
(b) Test the appropriate hypotheses and draw conclusions using the analysis of variance with α = 0.05
(c) Graphically analyze the interaction.
(d) Analyze the residuals from the experiment.
(e) Use Fisher's LSD method to investigate the differences between mean effects of insulation type. Use α =0.05.

Insulation	Temperature (°F)					
	1		2		3	
	6.6	4	4.5	2.2	2.3	0.9
	2.7	6.2	5.5	2.7	5.6	4.9
1	6	5	4.8	5.8	2.2	3.4
	3	3.2	3	1.5	1.3	3.3
	2.1	4.1	2.5	2.6	0.5	1.1
2	5.9	2.5	0.4	3.5	1.7	0.1
	5.7	4.4	8.9	7.7	2.6	9.9
	3.2	3.2	7	7.3	11.5	10.5
3	5.3	9.7	8	2.2	3.4	6.7
	7	8.9	12	9.7	8.3	8
	7.3	9	8.5	10.8	10.4	9.7
4	8.6	11.3	7.9	7.3	10.6	7.4

`(a) $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1,2,3 \\ j = 1,2,3,4 \\ k = 1,2,3,4,5,6 \end{cases}$

(b) $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ H_1 : at least one $\tau_j \neq 0$

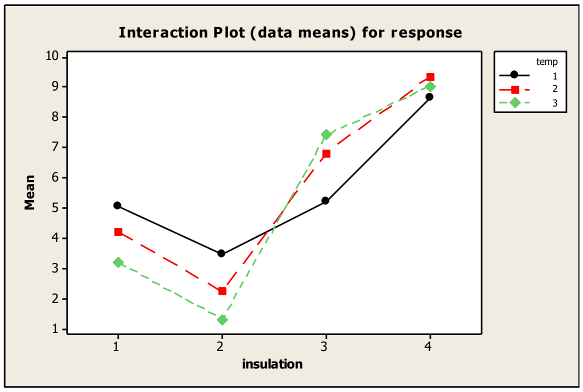
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ H_1 : at least one $\beta_j \neq 0$

$H_0 : (\tau\beta)_{11} = \ldots = (\tau\beta)_{ab} = 0$ H_1 : at least one $(\tau\beta)_{ij} \neq 0$

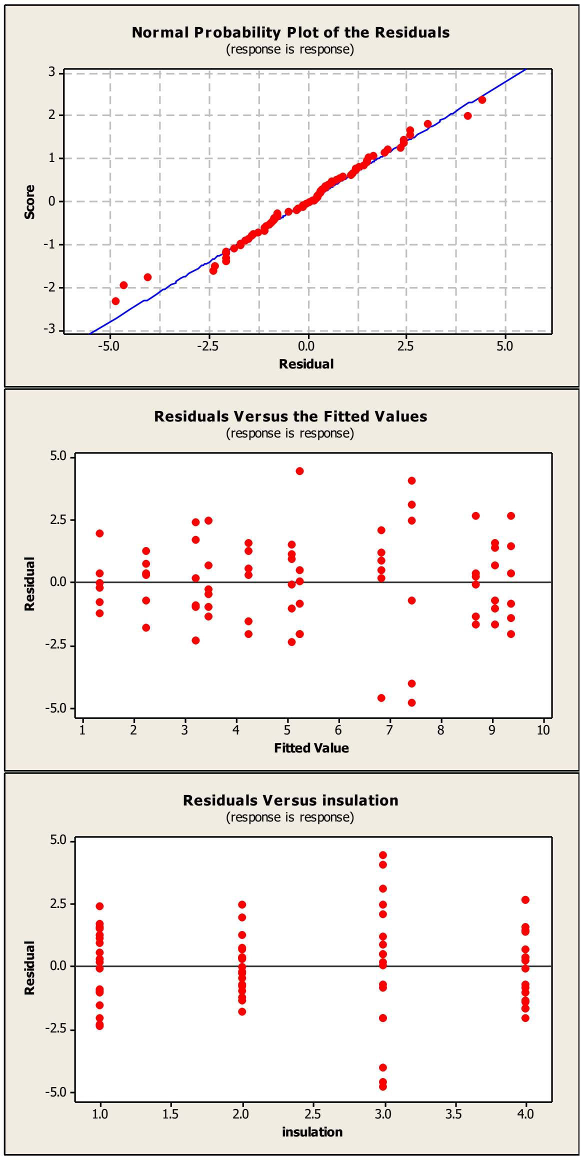
```
Source            DF       SS       MS      F      P
insulation         3  453.608  151.203  40.07  0.000
temp               2    2.443    1.222   0.32  0.725
insulation*temp    6   38.536    6.423   1.70  0.136
Error             60  226.432    3.774
Total             71  721.019

S = 1.94264   R-Sq = 68.60%   R-Sq(adj) = 62.84%
```

There is only one significant main effect, *insulation*. The temperature and interaction effects are not significant.

(c) Although there is some crossing of the lines, the interaction effect is minimal and was not found to be statistically significant in part (b).



(d) There is more variability for insulation type 3. The normality assumption is reasonable.



(e) Here, because only one of the main effects was significant, a model which included only insulation type was fit and LSD comparisons are made from that model:

```
Source       DF      SS      MS      F      P
insulation    3  453.61  151.20  38.45  0.000
Error        68  267.41    3.93
Total        71  721.02

S = 1.983   R-Sq = 62.91%   R-Sq(adj) = 61.28%


Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of insulation

Simultaneous confidence level = 80.02%

insulation = 1 subtracted from:

insulation   Lower  Center   Upper  ------+---------+---------+---------+---
2           -3.152  -1.833  -0.514            (-*--)
3            1.009   2.328   3.647                    (--*-)
4            3.537   4.856   6.175                         (--*-)
                                    ------+---------+---------+---------+---
                                       -5.0       0.0       5.0      10.0

insulation = 2 subtracted from:

insulation  Lower  Center  Upper  ------+---------+---------+---------+---
3           2.842   4.161  5.480                        (-*--)
4           5.370   6.689  8.008                             (-*--)
                                  ------+---------+---------+---------+---
                                     -5.0       0.0       5.0      10.0

insulation = 3 subtracted from:

insulation  Lower  Center  Upper  ------+---------+---------+---------+---
4           1.209   2.528  3.847                  (--*--)
                                  ------+---------+---------+---------+---
                                     -5.0       0.0       5.0      10.0
```

Because none of the intervals contain zero, all four insulation types are significantly different.

14-6 Johnson and Leone (*Statistics and Experimental Design in Engineering and the Physical Sciences,* John Wiley, 1977) described an experiment conducted to investigate warping of copper plates. The two factors studied were temperature and the copper content of the plates. The response variable is the amount of warping. The data are as follows:

Temperature (°C)	Copper Content (%)			
	40	60	80	100
50	17, 20	16, 21	24, 22	28, 27
75	12, 9	18, 13	17, 12	27, 31
100	16, 12	18, 21	25, 23	30, 23
125	21, 17	23, 21	23, 22	29, 31

(a) Is there any indication that either factor affects the amount of warping? Is there any interaction between the factors? Use $\alpha = 0.05$.
(b) Analyze the residuals from this experiment.
(c) Plot the average warping at each level of copper content and compare the levels using Fisher's LSD method. Describe the differences in the effects of the different levels of copper content on warping. If low warping is desirable, what level of copper content would you specify?
(d) Suppose that temperature cannot be easily controlled in the environment in which the copper plates are to be used.

Does this change your answer for part (c)?

(a)

1. $H_0 : \tau_1 = \tau_2 = 0$

   H_1 : at least one $\tau_j \neq 0$

2. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

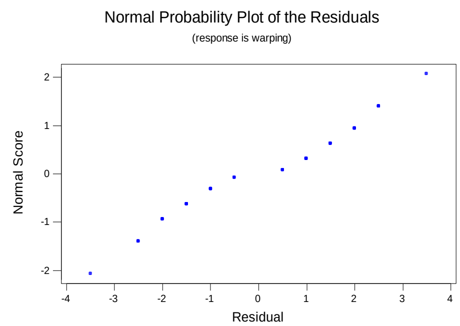
   H_1 : at least one $\beta_j \neq 0$

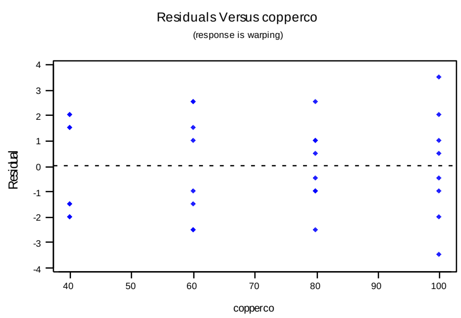
Analysis of Variance for warping

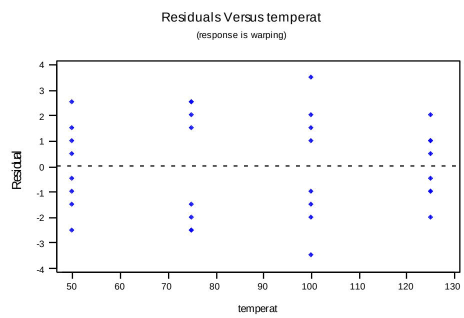
Source	DF	SS	MS	F	P
temp	3	156.09	52.03	7.67	0.002
copper	3	698.34	232.78	34.33	0.000
Interaction	9	113.78	12.64	1.86	0.133
Error	16	108.50	6.78		
Total	31	1076.72			

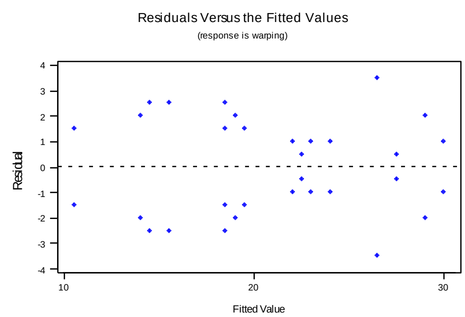
Reject H_0 for both of the main effects and conclude that both temperature and copper content have an effect on the mean warping. The interaction is not significant.

(b) The residuals for this experiment appear reasonable.

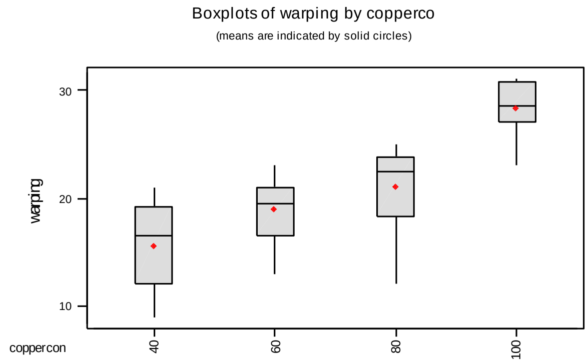








(c)




```
Fisher's pairwise comparisons
Family error rate = 0.195
Individual error rate = 0.0500
Critical value = 2.048
Intervals for (column level mean) - (row level mean)
                 40          60          80
      60      -7.139
               0.389
      80      -9.264      -5.889
              -1.736       1.639
     100     -16.514     -13.139     -11.014
              -8.986      -5.611      -3.486
```

There are significant differences in the following temperature levels:
40 and 80, 40 and 100,
60 and 100,
80 and 100
This difference is apparent on the boxplot and using Fisher's LSD method. If low warping is desired, temperature level 40 is most desirable.

(d) No, because the factors do not interact.

14-7 An article in the *IEEE Transactions on Electron Devices* (November 1986, p. 1754) described a study on the effects of two variables—polysilicon doping and anneal conditions (time and temperature)—on the base current of a bipolar transistor. The data from this experiment follow.

(a) Is there any evidence to support the claim that either polysilicon doping level or anneal conditions affect base current? Do these variables interact? Use α = 0.05.
(b) Graphically analyze the interaction.
(c) Analyze the residuals from this experiment.
(d) Use Fisher's LSD method to isolate the effects of anneal conditions on base current, with α = 0.05.

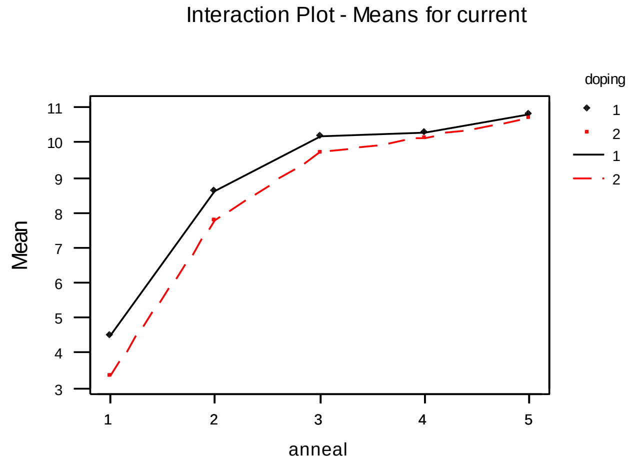
		Anneal (temperature/time)				
		900 60	900 180	950 60	1000 15	1000 30
Polysilicon doping	1×10^{20}	4.40	8.30	10.15	10.29	11.01
		4.60	8.90	10.20	10.30	10.58
	2×10^{20}	3.20	7.81	9.38	10.19	10.81
		3.50	7.75	10.02	10.10	10.60

(a)
```
Analysis of Variance for current
Source            DF          SS          MS         F        P
doping             1       1.442       1.442     25.23    0.000
anneal             4     124.238      31.059    543.52    0.000
doping*anneal      4       0.809       0.202      3.54    0.048
```

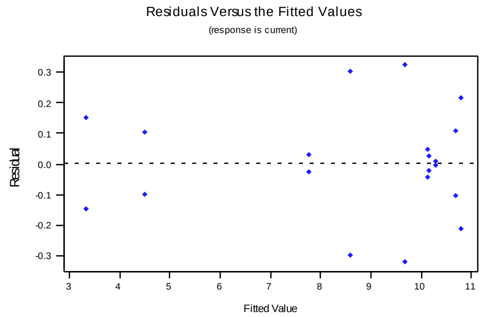
```
Error            10       0.571      0.057
Total            19     127.060
```

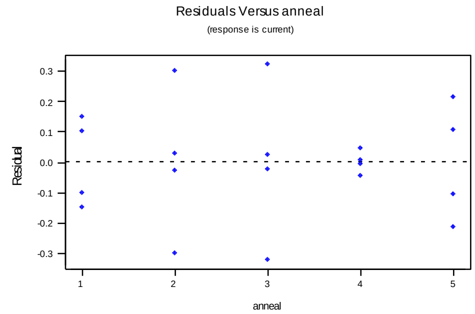
Both main factors are highly significant. The interaction is slightly significant at the 0.05 level of significance.

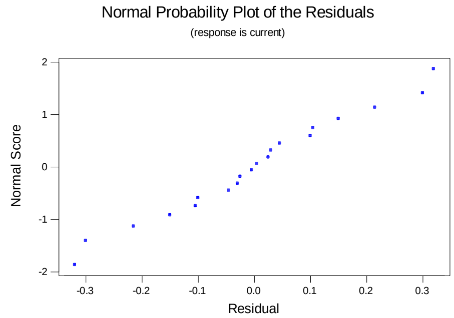
(b) The interaction plot shows that there is a slight interaction because the lines are not parallel.

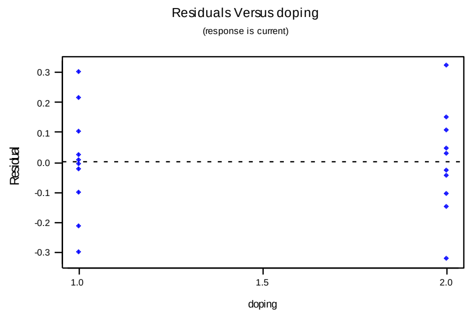


(c) Analysis of the residual plots shows that all there is no problem with the model adequacy or the assumptions necessary to build the model.









(d)

```
Fisher's pairwise comparisons
Family error rate = 0.258
Individual error rate = 0.0500
Critical value = 2.131
Intervals for (column level mean) - (row level mean)
                     1           2           3           4
          2    -4.9187
               -3.6113
          3    -6.6662     -2.4012
               -5.3588     -1.0938
          4    -6.9487     -2.6837     -0.9362
               -5.6413     -1.3763      0.3712
          5    -7.4787     -3.2137     -1.4662     -1.1837
               -6.1713     -1.9063     -0.1588      0.1237
```

There are significant differences in the annealing levels
1 and 2, 1 and 3, 1 and 4, 1 and 5,
2 and 3, 2 and 4, 2 and 5.
Therefore levels 1 and 2 are different from the other three.

14-8 An article in the *Journal of Testing and Evaluation* (1988, Vol. 16, pp. 508–515) investigated the effects of cyclic loading frequency and environment conditions on fatigue crack growth at a constant 22 MPa stress for a particular material. The data follow. The response variable is fatigue crack growth rate.

		Environment		
		Air	**H_2O**	**Salt H_2O**
Frequency	10	2.29 2.47 2.48 2.12	2.06 2.05 2.23 2.03	1.90 1.93 1.75 2.06
	1	2.65 2.68 2.06 2.38	3.20 3.18 3.96 3.64	3.10 3.24 3.98 3.24
	0.1	2.24 2.71 2.81 2.08	11.00 11.00 9.06 11.30	9.96 10.01 9.36 10.40

(a) Is there indication that either factor affects crack growth rate? Is there any indication of interaction? Use $\alpha = 0.05$.
(b) Analyze the residuals from this experiment.
(c) Repeat the analysis in part (a) using $ln(y)$ as the response. Analyze the residuals from this new response variable and comment on the results.

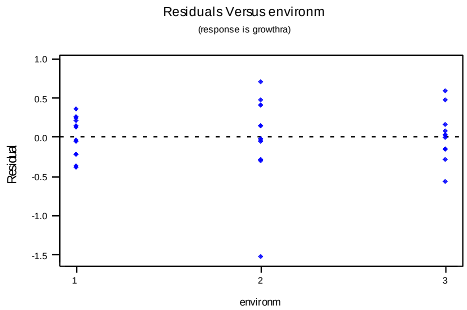
(a) Analysis of Variance for Crack Growth

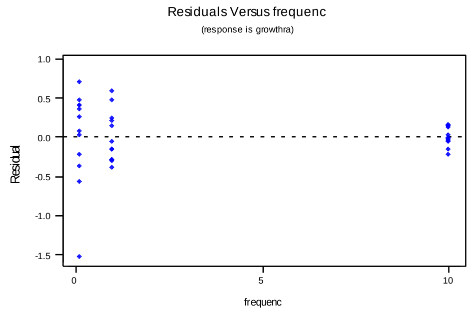
```
Source              DF        SS         MS        F        P
frequenc             2    209.893    104.946   522.40    0.000
environm             2     64.252     32.126   159.92    0.000
```

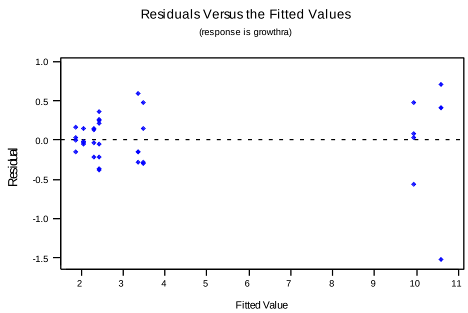
frequenc*environm	4	101.966	25.491	126.89	0.000
Error	27	5.424	0.201		
Total	35	381.535			

Both main factors and the interaction are significant.

(b) There appear to be some problems with constant variance in the residual plots.





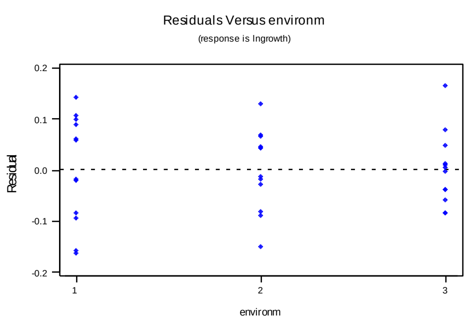


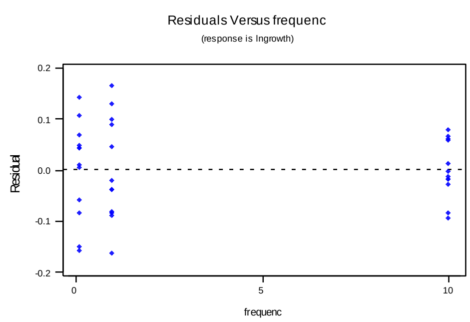
(c) Analysis of Variance of Ln(Crack Growth)

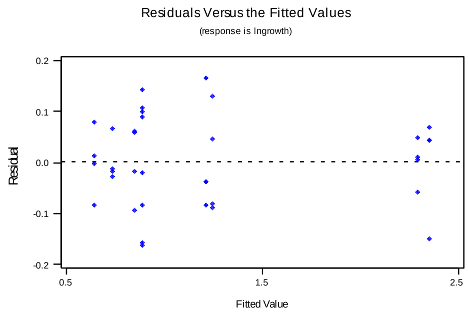
Source	DF	SS	MS	F	P
frequenc	2	7.5702	3.7851	404.09	0.000
environm	2	2.3576	1.1788	125.85	0.000
frequenc*environm	4	3.5284	0.8821	94.17	0.000
Error	27	0.2529	0.0094		
Total	35	13.7092			

The factors frequency, environment, and their interaction are all significant using the log of the data in the ANOVA

Residual plots on the log scale are improved. The variance appears to be more constant.







14-9 Consider a two-factor factorial experiment. Develop a formula for finding a 100(1 − α)% confidence interval on the difference between any two means for either a row or column factor. Apply this formula to find a 95% CI on the difference in mean warping at the levels of copper content 60 and 80% in Exercise 14-6.

The ratio $T = \dfrac{\bar{y}_{\cdot i\cdot} - \bar{y}_{\cdot j\cdot} - (\mu_i - \mu_j)}{\sqrt{2MS_E / n}}$ has a t distribution with $ab(n$-1) degrees of freedom

Therefore, the (1-α)% confidence interval on the difference in two treatment means is

$$\bar{y}_{\cdot i\cdot} - \bar{y}_{\cdot j\cdot} - t_{a/2,ab(n-1)}\sqrt{\frac{2MS_E}{n}} \le \mu_i - \mu_j \le \bar{y}_{\cdot i\cdot} - \bar{y}_{\cdot j\cdot} + t_{a/2,ab(n-1)}\sqrt{\frac{2MS_E}{n}}$$

The mean warping at 80% and 60% copper concentrations are 21.0 and 18.88, respectively. Here a = 4, b = 4, n = 2 and MS_E = 6.78. The degrees of freedom are (4)(4)(1) =16

$$(21.0 - 18.88) - 2.120\sqrt{\frac{2(6.78)}{2}} \le \mu_3 - \mu_2 \le (21.0 - 18.88) + 2.120\sqrt{\frac{2(6.78)}{2}}$$
$$-3.40 \le \mu_3 - \mu_2 \le 7.64$$

Therefore, there is no significant difference between the mean warping values at 80% and 60% copper concentration.

14-10 An article in *Journal of Chemical Technology and Biotechnology* [“A Study of Antifungal Antibiotic Production by Thermomonospora sp MTCC 3340 Using Full Factorial Design” (2003, Vol. 78, pp. 605-610)] considered the effects of several factors on antifungal activities. The antifungal yield was expressed as Nystatin international units per cm^3. The results from carbon source concentration (glucose) and incubation temperature factors follow.
See Table E14-1.

TABLE • E14-1 Data for Antifungal Activities

	Temperature (°C)								
Carbon(%)	25			30			37		
2	25.84	51.86	32.59	51.86	131.33	41.11	41.11	104.11	32.59
5	20.48	25.84	12.87	41.11	104.11	32.59	32.59	82.53	25.84
7.5	20.48	25.84	10.2	65.42	82.53	51.86	51.86	65.42	41.11

(a) State the hypotheses of interest.
(b) Test your hypotheses with α = 0.5.
(c) Analyze the residuals and plot the residuals versus the predicted yield.
(d) Using Fisher’s LSD method, compare the means of antifungal activity for the different carbon source concentrations.

(a) Hypotheses

$$H_0: \tau_1 = \tau_2 = \tau_3 = 0,\ \ H_1: at\ least\ one\ \tau_i \neq 0$$

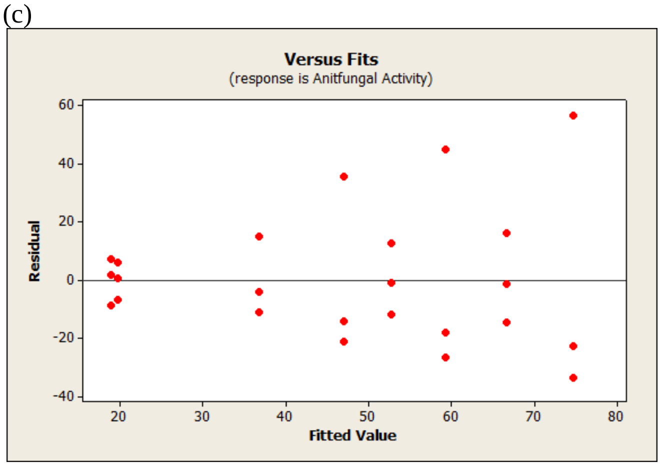
$H_0: \beta_1 = \beta_2 = \beta_3 = 0,\ \ H_1: at\ least\ one\ \beta_i \neq 0$

(b) ANOVA Table

```
Analysis of Variance for Anitfungal Activity, using Adjusted SS for Tests
Source               DF    Seq SS    Adj SS   Adj MS     F      P
Carbon                2    1072.8    1072.8    536.4  0.68  0.520
Temperature           2    8146.7    8146.7   4073.4  5.15  0.017
Carbon*Temperature    4     126.4     126.4     31.6  0.04  0.997
Error                18   14224.3   14224.3    790.2
Total                26   23570.3

S = 28.1112   R-Sq = 39.65%   R-Sq(adj) = 12.83%
```

The only P-value < 0 .05 is for *Temperature*. Therefore, only the main effect of *Temperature* is significant at α = 0.05.

(c)



The variance of the residuals increases as the fitted value increases.

(d) Here $t_{0.05/2,\,24}$ = 2.064 and the pooled standard deviation is 28.1112. Therefore, the standard error of the difference between two mean is $28.1112(1/9 + 1/9)^{1/2} = 13.252$.

```
Means

Carbon Level  N      Mean Yield
2.0           9          56.933
5.0           9          41.996
7.5           9          46.080
```

The largest difference in means is between Carbon 2.0 and Carbon 5.0 and this difference is 56.933 – 41.996 = 14.937. This is only slightly greater than the standard error of the difference. Therefore, there are no significant differences among the levels of carbon. This result agrees with the conclusions from the ANOVA.

14-11 An article in *Bioresource Technology* ["Quantitative Response of Cell Growth and Tuber Polysaccharides Biosynthesis by Medicinal Mushroom Chinese Truffl e Tuber Sinense to Metal Ion in Culture Medium" (2008, Vol. 99(16), pp. 7606–7615)] described an experiment to investigate the effect of metal ion concentration to the production of extracellular polysaccharides (EPS). It is suspected that Mg^{2+} and K^{+} (in millimolars) are related to EPS. The data from a full factorial design follow.

(a) State the hypotheses of interest.
(b) Test the hypotheses with α = 0.5.
(c) Analyze the residuals and plot residuals versus the predicted production.

Run	Mg^{2+} (mM)	K^+ (mM)	EPS (g/L)
1	40	5	3.88
2	50	15	4.23
3	40	10	4.67
4	30	5	5.86
5	50	10	4.50
6	50	5	3.62
7	30	15	3.84
8	40	15	3.25
9	30	10	4.18

(a)

$H_0: \tau_1 = \tau_2 = \tau_3 = 0,\ \ H_1: at\ least\ one\ \tau_i \neq 0$

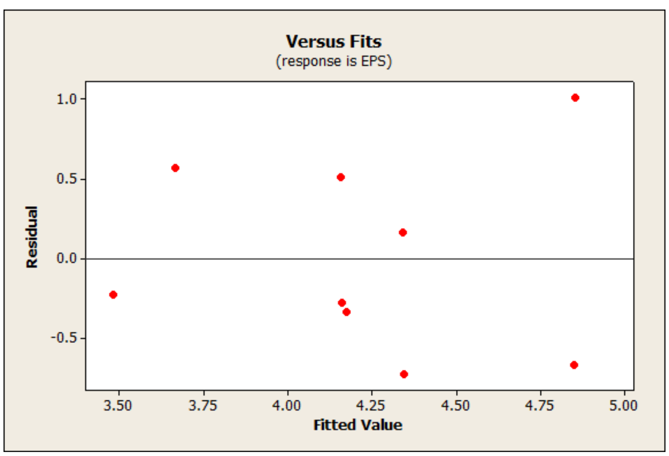
$H_0: \beta_1 = \beta_2 = \beta_3 = 0,\ \ H_1: at\ least\ one\ \beta_i \neq 0$

(b)

```
Analysis of Variance for EPS, using Adjusted SS for Tests

Source  DF  Seq SS  Adj SS  Adj MS     F      P
Mg+      2  0.7744  0.7744  0.3872  0.55  0.617
K+       2  0.9203  0.9203  0.4601  0.65  0.570
Error    4  2.8381  2.8381  0.7095
Total    8  4.5328


S = 0.842335   R-Sq = 37.39%   R-Sq(adj) = 0.00%
```

Neither main effect is significant. Because there are no replicates, the interaction between Mg+ and K+ is not modeled.

(c) From the residuals vs. fitted values plot, no departures from assumptions are evident.



Section 14-4

14-12 The quality control department of a fabric finishing plant is studying the effects of several factors on dyeing for a blended cotton/synthetic cloth used to manufacture shirts. Three operators, three cycle times, and two temperatures were selected, and three small specimens of cloth were dyed under each set of conditions. The finished cloth was compared to a standard, and a numerical score was assigned. The results are shown in the following table.

(a) State and test the appropriate hypotheses using the analysis of variance with $\alpha = 0.05$.

	Temperature					
	300°			350°		
	Operator			Operator		
Cycle Time	**1**	**2**	**3**	**1**	**2**	**3**
	23	27	31	24	38	34
40	24	28	32	23	36	36
	25	26	28	28	35	39
	36	34	33	37	34	34
50	35	38	34	39	38	36
	36	39	35	35	36	31
	28	35	26	26	36	28
60	24	35	27	29	37	26
	27	34	25	25	34	34

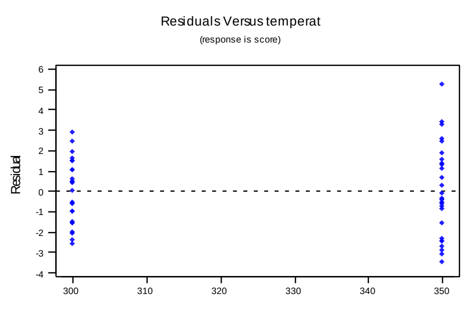
(b) The residuals may be obtained from $e_{ijkl} = y_{ijkl} - \bar{y}_{ijk\cdot}$. Graphically analyze the residuals from this experiment.

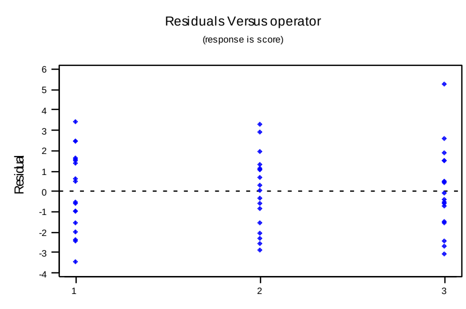
(a) `Analysis of Variance for dying score, using Adjusted SS for Tests`

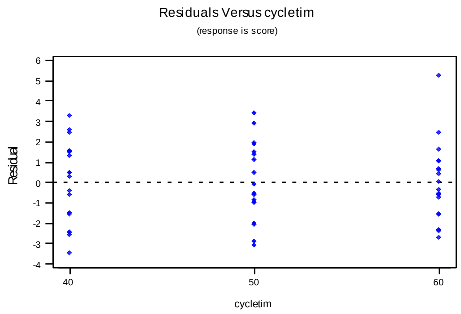
```
Source      DF        SS        MS      F      P
Time         2   396.778   198.389  39.85  0.000
Temp         1    73.500    73.500  14.77  0.000
Oper         2   256.333   128.167  25.75  0.000
Time*Temp    2    70.778    35.389   7.11  0.002
Time*Oper    4   300.222    75.056  15.08  0.000
Temp*Oper    2    14.111     7.056   1.42  0.254
Error       40   199.111     4.978
Total       53  1310.833

S = 2.23109   R-Sq = 84.81%   R-Sq(adj) = 79.87%
```

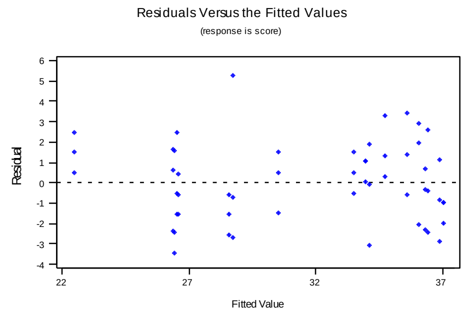
Only the operator*temperature interaction is not significant.

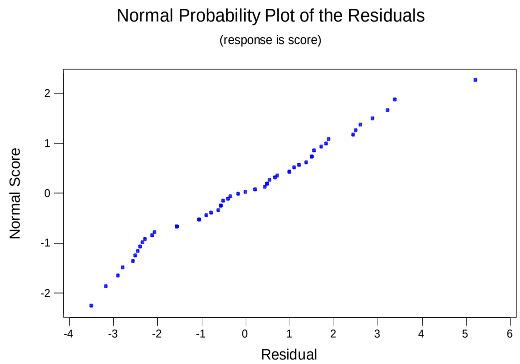
(b)











The residuals are acceptable.

14-13 The percentage of hardwood concentration in raw pulp, the freeness, and the cooking time of the pulp are being investigated for their effects on the strength of paper. The data from a three-factor factorial experiment are shown in the following table.

(a) Analyze the data using the analysis of variance assuming that all factors are fixed. Use α = 0.05.
(b) Compute approximate *P*-values for the *F*-ratios in part (a).
(c) The residuals are found from $e_{ijkl} = y_{ijkl} - \bar{y}_{ijk\cdot}$. Graphically analyze the residuals from this experiment.

Hardwood Concentration %	Cooking Time 1.5 hours			Cooking Time 2.0 hours		
	Freeness			Freeness		
	350	500	650	350	500	650
	96.6	97.7	99.4	98.4	99.6	100.6
10	96.0	96.0	99.8	98.6	100.4	100.9
	98.5	96.0	98.4	97.5	98.7	99.6
15	97.2	96.9	97.6	98.1	96.0	99.0
	97.5	95.6	97.4	97.6	97.0	98.5
20	96.6	96.2	98.1	98.4	97.8	99.8

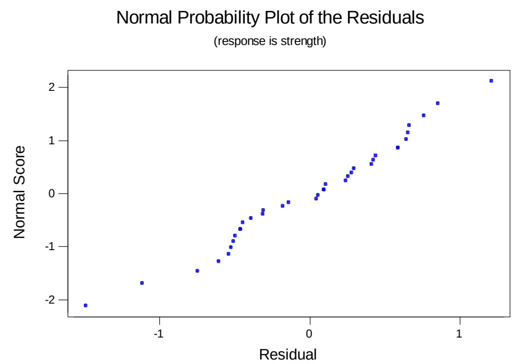
Parts (a) and (b)

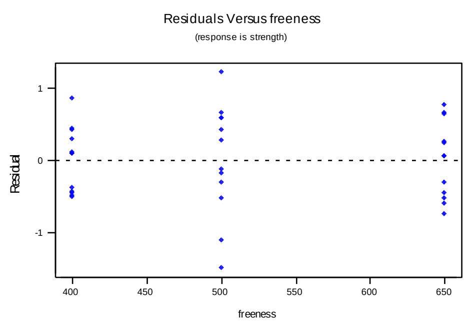
```
Analysis of Variance for strength
  Source              DF        SS        MS       F      P
  hardwood             2    8.3750    4.1875    7.64  0.003
  cookingtime          1   17.3611   17.3611   31.66  0.000
  freeness             2   21.8517   10.9258   19.92  0.000
```

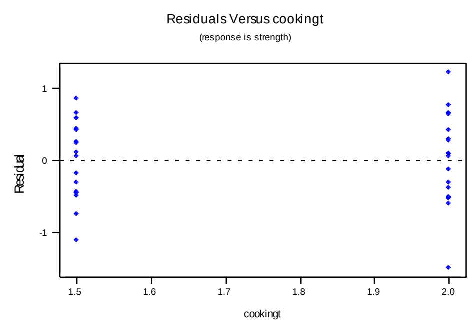
hardwood*cookingtime	2	3.2039	1.6019	2.92	0.075
hardwood*freeness	4	6.5133	1.6283	2.97	0.042
cookingtime*freeness	2	1.0506	0.5253	0.96	0.399
Error	22	12.0644	0.5484		
Total	35	70.4200			

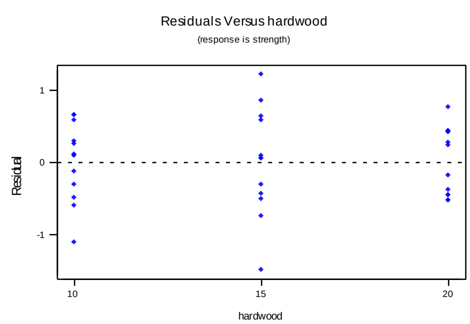
All main factors are significant. The interaction of hardwood*freeness is also significant.

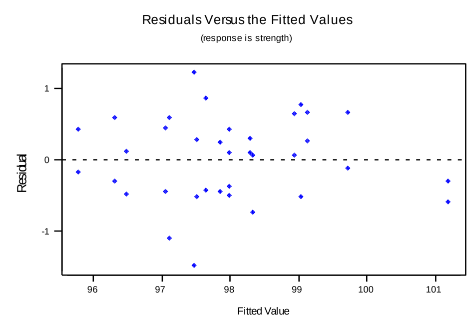
(c) The residual plots do not indicate serious problems with normality or equality of variance.











Section 14-5

14-14 An engineer is interested in the effect of cutting speed (*A*), metal hardness (*B*), and cutting angle (*C*) on the life of a cutting tool. Two levels of each factor are chosen, and two replicates of a 2^3 factorial design are run. The tool life data (in hours) are shown in the following table.

Treatment Combination	Replicate	
	I	II
(1)	221	311
a	325	435
b	354	348
ab	552	472
c	440	453
ac	406	377
bc	605	500
abc	392	419

(a) Analyze the data from this experiment.
(b) Find an appropriate regression model that explains tool life in terms of the variables used in the experiment.
(c) Analyze the residuals from this experiment.

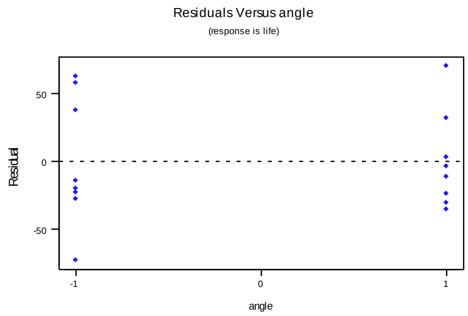
(a)
```
Analysis of Variance for life (coded units)

Source   DF      SS      MS      F      P
A         1    1332    1332   0.54  0.483
B         1   28392   28392  11.53  0.009
C         1   20592   20592   8.36  0.020
A*B       1     506     506   0.21  0.662
A*C       1   56882   56882  23.10  0.001
B*C       1    2352    2352   0.96  0.357
A*B*C     1    4830    4830   1.96  0.199
Error     8   19700    2463
Total    15  134588

S = 49.6236   R-Sq = 85.36%   R-Sq(adj) = 72.56%
```

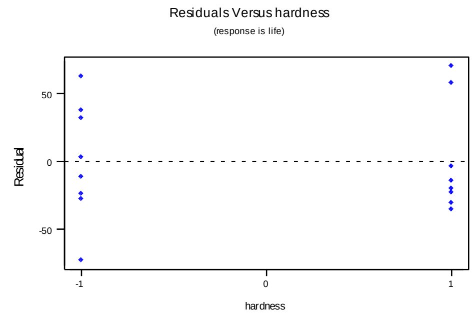
Hardness, angle and the speed*angle interaction are significant.

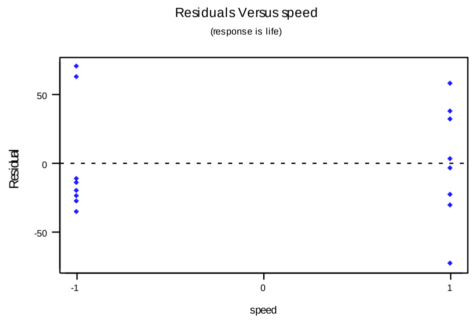
(b)
```
Estimated Effects and Coefficients for life (coded units)

Term                    Effect     Coef    SE Coef      T      P
Constant                         413.13      12.41  33.30  0.000
speed                    18.25     9.12      12.41   0.74  0.483
hardness                 84.25    42.12      12.41   3.40  0.009
angle                    71.75    35.87      12.41   2.89  0.020
speed*hardness          -11.25    -5.63      12.41  -0.45  0.662
speed*angle            -119.25   -59.62      12.41  -4.81  0.001
hardness*angle          -24.25   -12.12      12.41  -0.98  0.357
speed*hardness*angle    -34.75   -17.37      12.41  -1.40  0.199
```

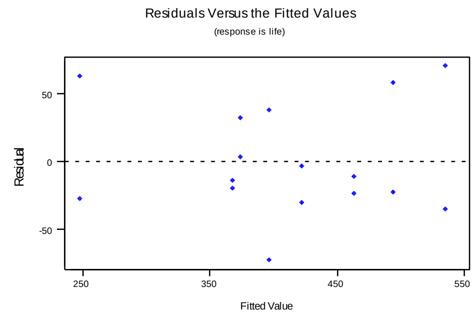
$\hat{y} = 413.125 + 9.125x_1 + 45.12x_2 + 35.87x_3 - 59.62x_1x_3$

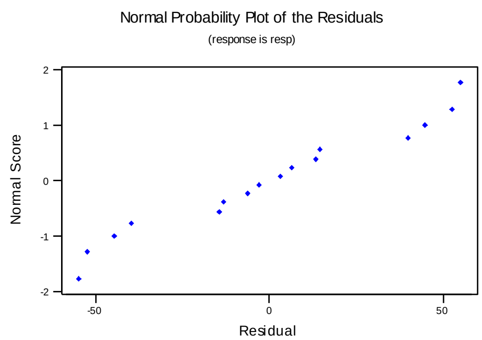
(c) Analysis of the residuals shows that all assumptions are reasonable.











14-15 Four factors are thought to influence the taste of a soft-drink beverage: type of sweetener (A), ratio of syrup to water (B), carbonation level (C), and temperature (D). Each factor can be run at two levels, producing a 2^4 design. At each run in the design, samples of the beverage are given to a test panel consisting of 20 people. Each tester assigns the beverage a point score from 1 to 10. Total score is the response variable, and the objective is to find a formulation that maximizes total score. Two replicates of this design are run, and the results are shown in the table. Analyze the data and draw conclusions. Use a = 0.05 in the statistical tests.

Treatment Combination	Replicate I	Replicate II
(1)	159	163
a	168	175
b	158	163
ab	166	168
c	175	178
ac	179	183
bc	173	168
abc	179	182
d	164	159
ad	187	189
bd	163	159
abd	185	191
cd	168	174
acd	197	199
bcd	170	174
abcd	194	198

```
Term          Effect      Coef    SE Coef       T      P
Constant               175.250     0.5467  320.59  0.000
A             17.000     8.500     0.5467   15.55  0.000
B             -1.625    -0.812     0.5467   -1.49  0.157
C             10.875     5.438     0.5467    9.95  0.000
D              8.375     4.187     0.5467    7.66  0.000
A*B           -0.125    -0.063     0.5467   -0.11  0.910
A*C           -0.625    -0.313     0.5467   -0.57  0.575
A*D            9.125     4.562     0.5467    8.35  0.000
B*C           -0.250    -0.125     0.5467   -0.23  0.822
B*D            1.250     0.625     0.5467    1.14  0.270
C*D           -1.250    -0.625     0.5467   -1.14  0.270
A*B*C          0.750     0.375     0.5467    0.69  0.503
A*B*D         -0.500    -0.250     0.5467   -0.46  0.654
A*C*D         -0.000    -0.000     0.5467   -0.00  1.000
B*C*D          0.125     0.063     0.5467    0.11  0.910
A*B*C*D       -1.625    -0.812     0.5467   -1.49  0.157
```

Factors A, C, and D are significant as well as the interaction AD.

14-16 The following data represent a single replicate of a 25 design that is used in an experiment to study the compressive strength of concrete. The factors are mix (*A*), time (*B*), laboratory (*C*), temperature (*D*), and drying time (*E*).

(1)	=	700	*e*	=	800
a	=	900	*ae*	=	1200
b	=	3400	*be*	=	3500
ab	=	5500	*abe*	=	6200
c	=	600	*ce*	=	600
ac	=	1000	*ace*	=	1200
bc	=	3000	*bce*	=	3006
abc	=	5300	*abce*	=	5500
d	=	1000	*de*	=	1900
ad	=	1100	*ade*	=	1500
bd	=	3000	*bde*	=	4000
abd	=	6100	*abde*	=	6500
cd	=	800	*cde*	=	1500
acd	=	1100	*acde*	=	2000
bcd	=	3300	*bcde*	=	3400

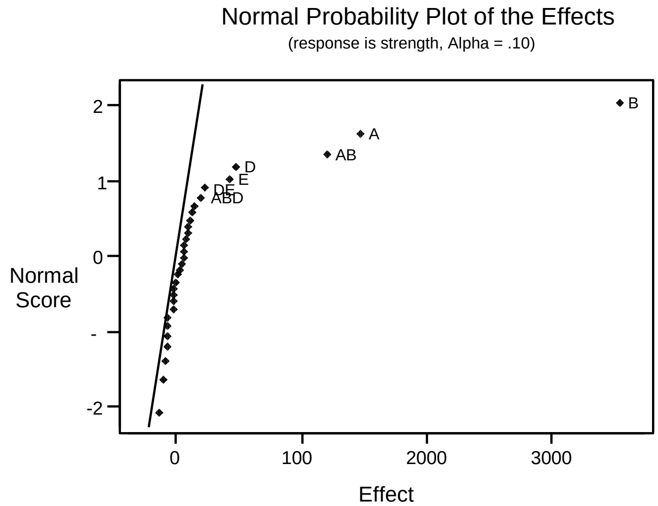
abcd	=	6000	*abcde*	=	6800

(a) Estimate the factor effects.
(b) Which effects appear important? Use a normal probability plot.
(c) If it is desirable to maximize the strength, in which direction would you adjust the process variables?
(d) Analyze the residuals from this experiment.

(a)

```
Estimated Effects and Coefficients for strength
Term              Effect
A                1462.13
B                3537.87
C                -137.12
D                 474.62
E                 425.38
A*B              1199.62
A*C               124.62
A*D                62.87
A*E                62.12
B*C               -99.63
B*D               -12.88
B*E               -12.12
C*D               112.13
C*E               -62.13
D*E               224.62
A*B*C             -62.88
A*B*D             200.38
A*B*E              49.63
A*C*D              75.38
A*C*E              99.63
A*D*E             -87.12
B*C*D              99.62
B*C*E             -74.63
B*D*E             -62.88
C*D*E              37.13
A*B*C*D           -12.12
A*B*C*E            12.13
A*B*D*E             0.37
A*C*D*E           150.37
B*C*D*E           -25.38
A*B*C*D*E          62.88
```

(b)



From the plot, the effects that appear to be important are A, B, D, E, and the interactions AB, DE, and ABD.

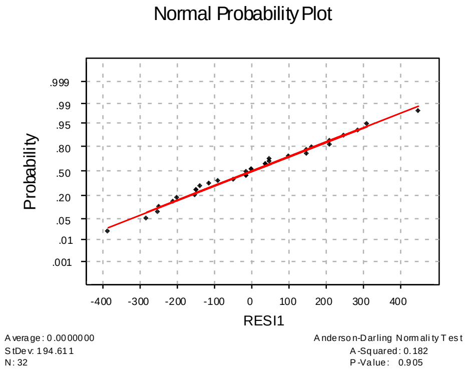
(c) To maximize strength, the variables A, B, D, and E should be increased. Variable C is not significant. Thus, any level of C is acceptable.

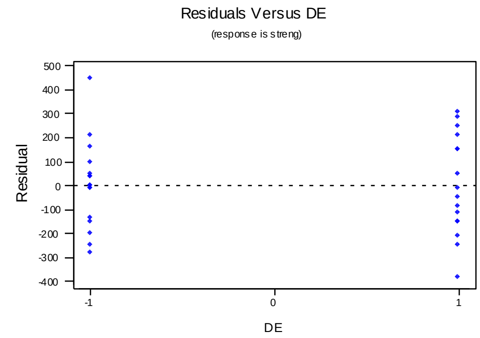
The regression equation is

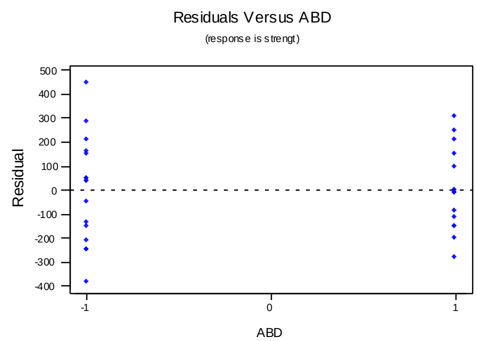
$$\hat{y} = 2888.7 + 731.1x_1 + 1768.9x_2 + 237.3x_4 + 212.7x_5 + 599.8x_1x_2 + 112.3x_4x_5 + 100.2x_1x_2x_4$$

```
Predictor       Coef       StDev          T        P
Constant     2887.69       39.10      73.86    0.000
A             731.06       39.10      18.70    0.000
B            1768.94       39.10      45.24    0.000
D             237.31       39.10       6.07    0.000
E             212.69       39.10       5.44    0.000
AB            599.81       39.10      15.34    0.000
DE            112.31       39.10       2.87    0.008
ABD           100.19       39.10       2.56    0.017
S = 221.2       R-Sq = 99.1%     R-Sq(adj) = 98.9%

Analysis of Variance
Source        DF          SS          MS          F        P
Regression     7   132722308    18960330     387.58    0.000
Error         24     1174077       48920
Total         31   133896385

Source        DF      Seq SS
A              1    17102476
B              1   100132476
D              1     1802151
E              1     1447551
AB             1    11512801
DE             1      403651
ABD            1      321201
```

(d)







The normal probability plot of the residuals indicates the assumption of normality is reasonable.
The model appears to be adequate.

14-17 An article in *IEEE Transactions on Semiconductor Manufacturing* (1992, Vol. 5, pp. 214–222) described an experiment to investigate the surface charge on a silicon wafer. The factors thought to influence induced surface charge are cleaning method (spin rinse dry or SRD and spin dry or SD) and the position on the wafer where the charge was measured. The surface charge ($\times 10^{11}$ q / cm3) response data follow:

		Test Position	
		L	**R**
Cleaning Method	**SD**	1.66	1.84
		1.90	1.84
		1.92	1.62
	SRD	–4.21	–7.58
		–1.35	–2.20
		–2.08	–5.36

(a) Estimate the factor effects.
(b) Which factors appear important? Use $\alpha = 0.05$.
(c) Analyze the residuals from this experiment.

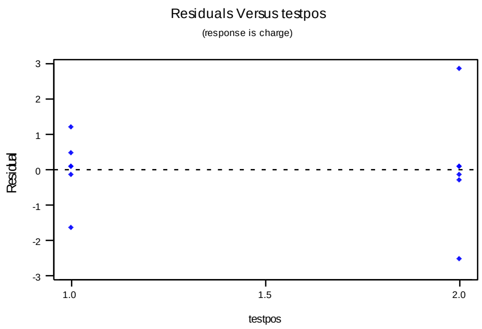
(a)
```
Estimated Effects and Coefficients for charge
Term                 Effect      Coef  StDev Coef       T      P
Constant                       -1.000      0.4462   -2.24  0.055
cleanmet             -5.593    -2.797      0.4462   -6.27  0.000
testpos              -1.280    -0.640      0.4462   -1.43  0.189
cleanmet*testpos     -1.220    -0.610      0.4462   -1.37  0.209
```

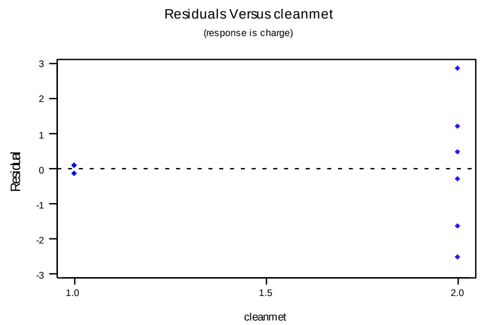
```
b)Analysis of Variance for charge
Source                 DF      Seq SS      Adj SS      Adj MS       F       P
```

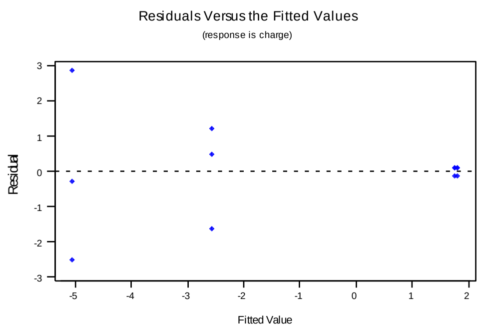
Main Effects	2	98.771	98.7713	49.386	20.67	0.001
2-Way Interactions	1	4.465	4.4652	4.465	1.87	0.209
Residual Error	8	19.110	19.1101	2.389		
Pure Error	8	19.110	19.1101	2.389		
Total	11	122.347				

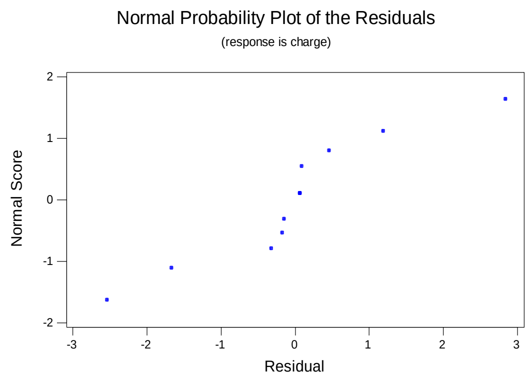
(b) Cleaning Method is the only significant factor.

(c) Analysis of the residuals shows that there is more variability at test position R and cleaning material SRD. In the case of the cleaning material, the difference in the variances is very large. The variation decreases with increased fitted values. The normal probability plot appears to have some variations from the straight line.









14-18 An article in *Oikos: A Journal of Ecology* ["Regulation of Root Vole Population Dynamics by Food Supply and Predation: A Two-Factor Experiment" (2005, Vol. 109, pp. 387–395)] investigated how food supply interacts with predation in the regulation of root vole (Microtus oeconomus Pallas) population dynamics. A replicated two-factor field experiment manipulating both food supply and predation condition for root voles was conducted. Four treatments were applied: −P, +F (no-predator, food-supplemented); +P, +F (predator-access, food-supplemented); −P, −F (no-predator, nonsupplemented); +P, −F (predator-access, food-supplemented). The population density of root voles (voles ha $^{-1}$) for each treatment combination follows.

Food Supply (F)	Predation (P)	Replicates		
+1	−1	88.589	114.059	200.979
+1	+1	56.949	97.079	78.759
−1	−1	65.439	89.089	172.339
−1	+1	40.799	47.959	74.439

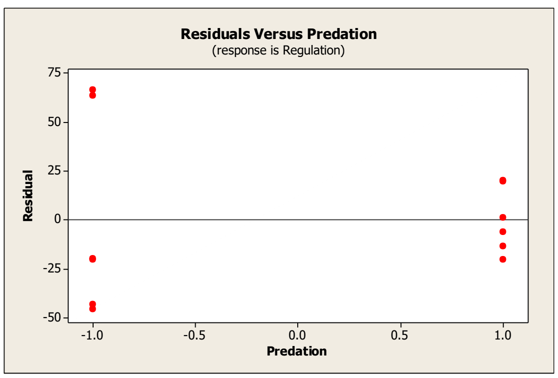
(a) What is an appropriate statistical model for this experiment?
(b) Analyze the data and draw conclusions.
(c) Analyze the residuals from this experiment. Are there any problems with model adequacy?

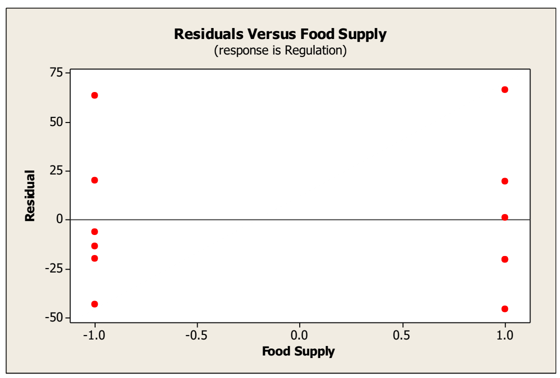
(a) $$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1,2 \\ j = 1,2 \\ k = 1,2,3 \end{cases}$$

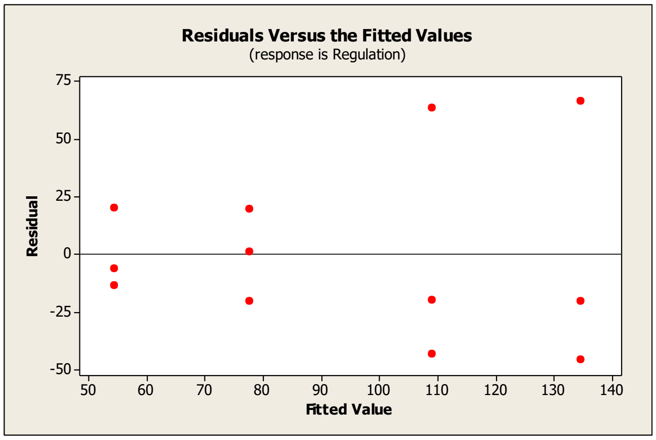
(b) There is no significant effect in the model.

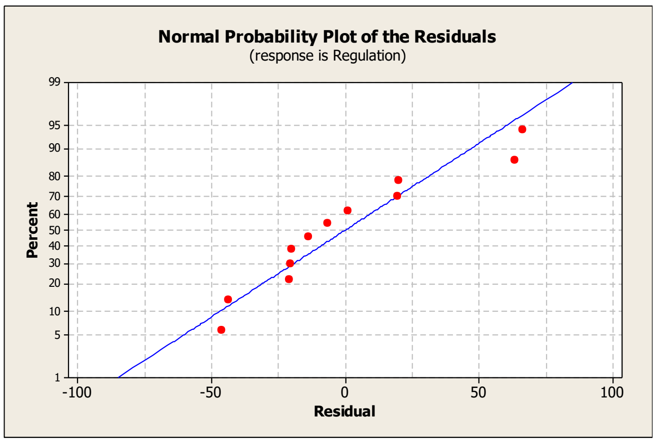
```
Analysis of Variance for Regulation (coded units)
Source            DF       SS       MS     F      P
Food Sup           1   1784.9  1784.86  0.97  0.353
Predatio           1   9324.7  9324.75  5.08  0.054
Interaction        1      4.3     4.28  0.00  0.963
Error              8  14686.2  1835.78
Total             11  25800.1
```

(c) There appear to be some problems with constant variance in the residual plots. The normal probability plot shows that the assumption of normality is reasonable.









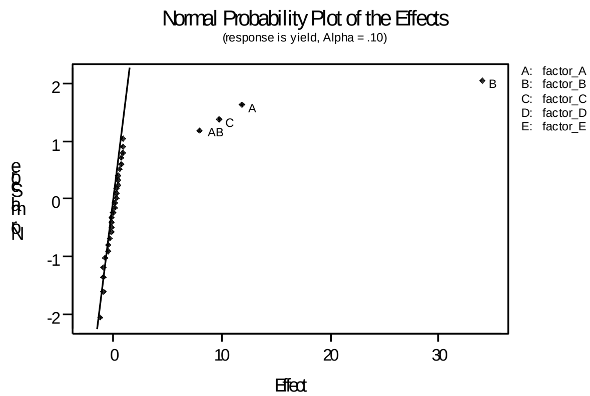
14-19 An experiment was run in a semiconductor fabrication plant in an effort to increase yield. Five factors, each at two levels, were studied. The factors (and levels) were *A* = aperture setting (small, large), *B* = exposure time (20% below nominal, 20% above nominal), *C* = development time (30 and 45 seconds), *D* = mask dimension (small, large), and *E* = etch time (14.5 and 15.5 minutes). The following unreplicated 2^5 design was run:

(1)	=	7	*e*	=	8
a	=	9	*ae*	=	12
b	=	34	*be*	=	35
ab	=	55	*abe*	=	52
c	=	16	*ce*	=	15
ac	=	20	*ace*	=	22
bc	=	40	*bce*	=	45
abc	=	60	*abce*	=	65
d	=	8	*de*	=	6
ad	=	10	*ade*	=	10
bd	=	32	*bde*	=	30
abd	=	50	*abde*	=	53
cd	=	18	*cde*	=	15
acd	=	21	*acde*	=	20
bcd	=	44	*bcde*	=	41
abcd	=	61	*abcde*	=	63

(a) Construct a normal probability plot of the effect estimates. Which effects appear to be large?
(b) Conduct an analysis of variance to confirm your findings for part (a).
(c) Construct a normal probability plot of the residuals. Is the plot satisfactory?
(d) Plot the residuals versus the predicted yields and versus each of the five factors. Comment on the plots.
(e) Interpret any significant interactions.
(f) What are your recommendations regarding process operating conditions?

(g) Project the 2^5 design in this problem into a 2^r for $r < 5$ de sign in the important factors. Sketch the design and show the average and range of yields at each run. Does this sketch aid in data interpretation?

(a) From the normal probability plot of the effects, factors A, B, C, and the AB interaction appear to be significant.

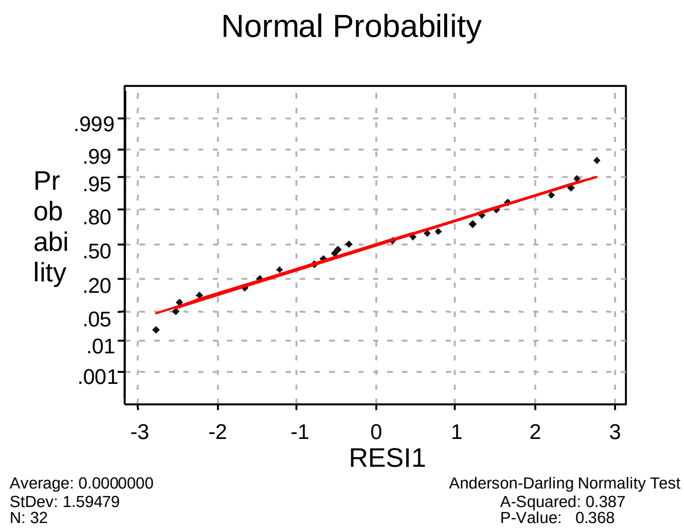


(b)

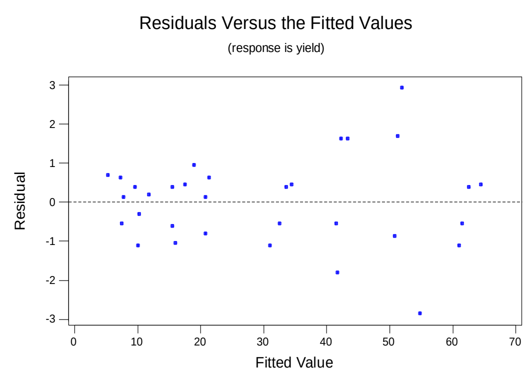
```
Analysis of Variance for yield (coded units)
Term                   Effect      Coef  StDev Coef        T      P
Constant                        30.5312      0.2786   109.57  0.000
factor_A              11.8125    5.9063      0.2786    21.20  0.000
factor_B              33.9375   16.9687      0.2786    60.90  0.000
factor_C               9.6875    4.8437      0.2786    17.38  0.000
factor_D              -0.8125   -0.4063      0.2786    -1.46  0.164
factor_E               0.4375    0.2187      0.2786     0.79  0.444
factor_A*factor_B      7.9375    3.9687      0.2786    14.24  0.000
factor_A*factor_C      0.4375    0.2187      0.2786     0.79  0.444
factor_A*factor_D     -0.0625   -0.0313      0.2786    -0.11  0.912
factor_A*factor_E      0.9375    0.4687      0.2786     1.68  0.112
factor_B*factor_C      0.0625    0.0313      0.2786     0.11  0.912
factor_B*factor_D     -0.6875   -0.3437      0.2786    -1.23  0.235
factor_B*factor_E      0.5625    0.2813      0.2786     1.01  0.328
factor_C*factor_D      0.8125    0.4062      0.2786     1.46  0.164
factor_C*factor_E      0.3125    0.1563      0.2786     0.56  0.583
factor_D*factor_E     -1.1875   -0.5938      0.2786    -2.13  0.049

Analysis of Variance for yield
Source                 DF      Seq SS      Adj SS      Adj MS       F       P
Main Effects            5     11087.9     11087.9     2217.58  892.61   0.000
2-Way Interactions     10       536.3       536.3       53.63   21.59   0.000
Residual Error         16        39.7        39.7        2.48
Total                  31     11664.0
```

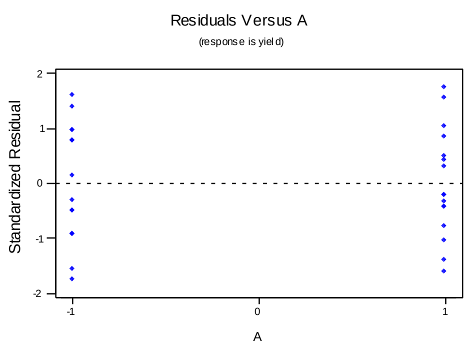
The analysis confirms our findings from part a)

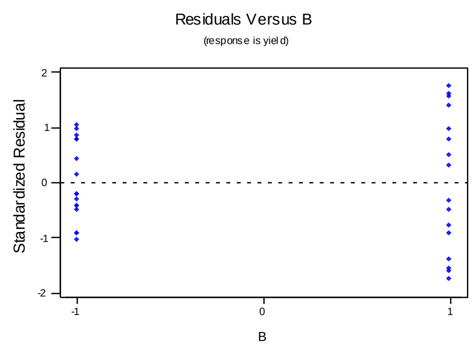
(c) The normal probability plot of the residuals is satisfactory. However residual variance appears to increase as the fitted value increases.

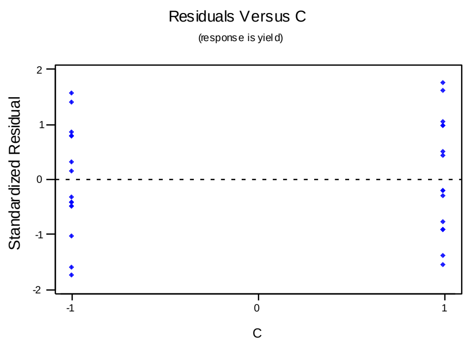


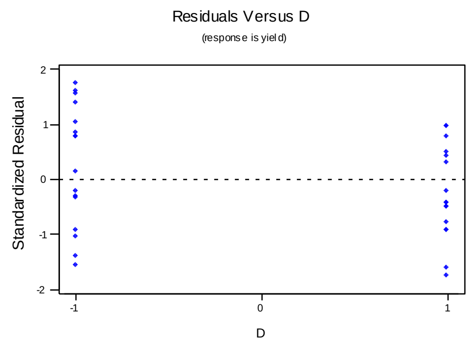
(d) All plots support the constant variance assumption, although there is a very slight indication that variability is greater at the high level of factor B.

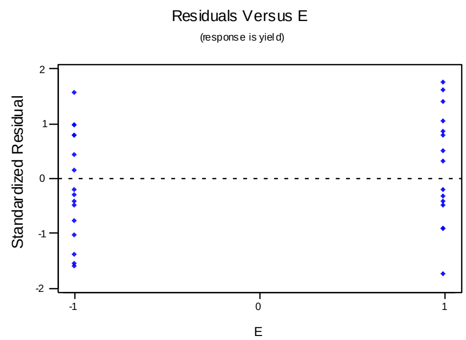




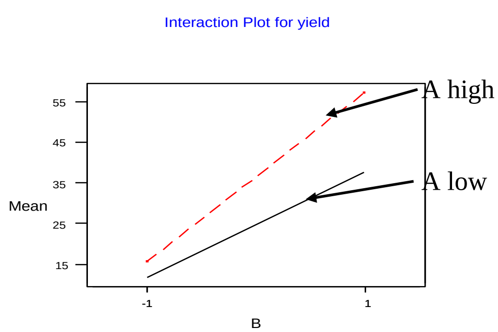






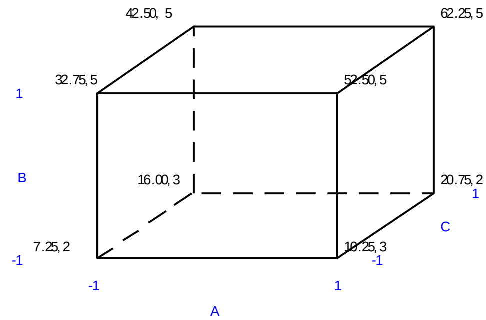


(e) The AB interaction appears to be significant. The interaction plot indicates that a high level of A and of B increases the mean yield, while low levels of both factors would lead to a reduction in the mean yield.



(f) To increase yield and optimize the process, we would want to set A, B, and C at their high levels.

(g) It is evident from the cube plot that we should run the process with all factors set at their high levels.

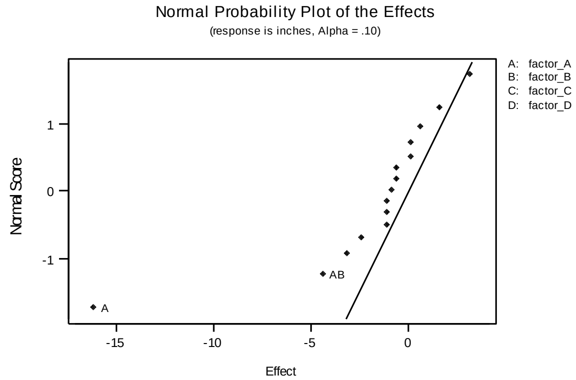


14-20 An experiment described by M. G. Natrella in the National Bureau of Standards' *Handbook of Experimental Statistics* (1963, No. 91) involves flame-testing fabrics after applying fire-retardant treatments. The four factors considered are type of fabric (A), type of fire-retardant treatment (B), laundering condition (C—the low level is no laundering, the high level is after one laundering), and method of conducting the flame test (D). All factors are run at two levels, and the response variable is the inches of fabric burned on a standard size test sample. The data are:

(1)	=	42	*d*	=	40
a	=	31	*ad*	=	30
b	=	45	*bd*	=	50
ab	=	29	*abd*	=	25
c	=	39	*cd*	=	40
ac	=	28	*acd*	=	25
bc	=	46	*bcd*	=	50
abc	=	32	*abcd*	=	23

(a) Estimate the effects and prepare a normal plot of the effects.
(b) Construct an analysis of variance table based on the model tentatively identified in part (a).
(c) Construct a normal probability plot of the residuals and comment on the results.

(a)




```
Estimated Effects and Coefficients for inches
Term                            Effect      Coef
Constant                                  35.938
factor_A                       -16.125    -8.063
factor_B                         3.125     1.562
factor_C                        -1.125    -0.562
factor_D                        -1.125    -0.562
factor_A*factor_B               -4.375    -2.187
factor_A*factor_C               -0.625    -0.312
```

```
factor_A*factor_D                  -3.125    -1.562
factor_B*factor_C                   1.625     0.812
factor_B*factor_D                   0.125     0.063
factor_C*factor_D                  -0.625    -0.313
factor_A*factor_B*factor_C          0.625     0.313
factor_A*factor_B*factor_D         -2.375    -1.188
factor_A*factor_C*factor_D         -1.125    -0.563
factor_B*factor_C*factor_D         -0.875    -0.437
factor_A*factor_B*factor_C
*factor_D                           0.125     0.063
```

According to the normal probability plot, factors A, B, and AB appear to be significant.

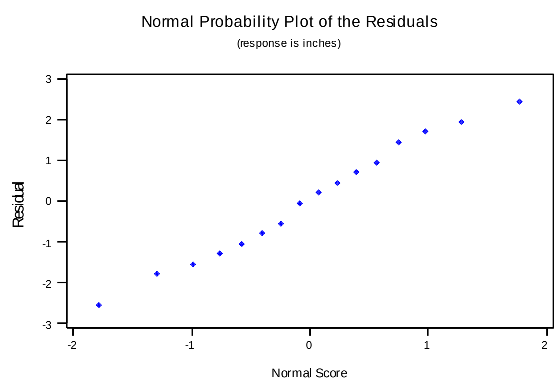
Parts (b) and (c)

Remove the three- and four-factor interactions to generate the following analysis:

Term	Effect	Coef	StDev Coef	T	P
Constant		35.938	0.6355	56.55	0.000
factor_A	-16.125	-8.063	0.6355	-12.69	0.000
factor_B	3.125	1.562	0.6355	2.46	0.057
factor_C	-1.125	-0.562	0.6355	-0.89	0.417
factor_D	-1.125	-0.562	0.6355	-0.89	0.417
factor_A*factor_B	-4.375	-2.187	0.6355	-3.44	0.018
factor_A*factor_C	-0.625	-0.312	0.6355	-0.49	0.644
factor_A*factor_D	-3.125	-1.562	0.6355	-2.46	0.057
factor_B*factor_C	1.625	0.812	0.6355	1.28	0.257
factor_B*factor_D	0.125	0.063	0.6355	0.10	0.925
factor_C*factor_D	-0.625	-0.313	0.6355	-0.49	0.644

Analysis of Variance for resp, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
A	1	1040.06	1040.06	1040.06	131.03	0.000
B	1	39.06	39.06	39.06	4.92	0.047
A*B	1	76.56	76.56	76.56	9.65	0.009
Error	12	95.25	95.25	7.94		
Total	15	1250.94				



14-21 Consider the data from Exercise 14-14. Suppose that the data from the second replicate were not available. Analyze the data from replicate I only and comment on your findings.

With only one replicate, the full factorial cannot be analyzed without using the 3-way interaction for error.

Estimated Effects and Coefficients for resp (coded units)

```
Term            Effect      Coef      SE Coef       T      P
Constant                  411.88        34.13   12.07  0.053
A                13.75      6.88        34.12    0.20  0.873
B               127.75     63.87        34.12    1.87  0.312
C                97.75     48.88        34.13    1.43  0.388
A*B             -21.25    -10.63        34.13   -0.31  0.808
A*C            -137.25    -68.63        34.13   -2.01  0.294
B*C             -52.25    -26.13        34.13   -0.77  0.584
Analysis of Variance for resp (coded units)

Source                DF      Seq SS     Adj SS     Adj MS      F      P
Main Effects           3       52128      52128      17376   1.87  0.483
2-Way Interactions     3       44038      44038      14679   1.58  0.516
Residual Error         1        9316       9316       9316
Total                  7      105483
```

The results do not indicate any significant effects.

14-22 A 2^4 factorial design was run in a chemical process. The design factors are A = time, B = concentration, C = pressure, and D = temperature. The response variable is yield. The data follow:

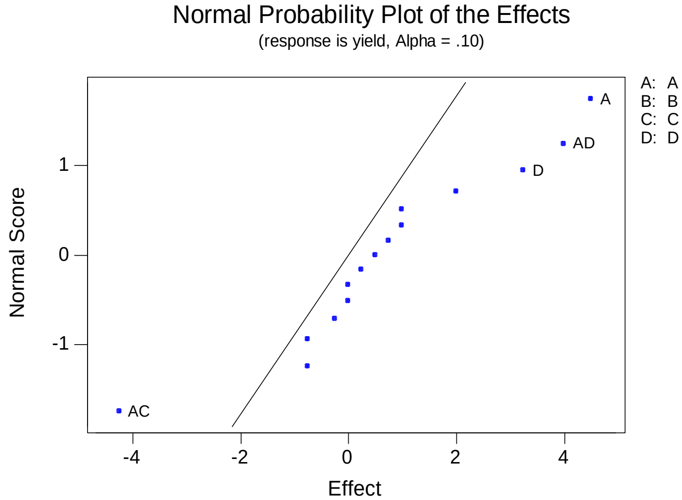
Run	A	B	C	D	Yield (pounds)	Factor Levels	–	+
1	–	–	–	–	12	A (hours)	2	3
2	+	–	–	–	18	B (%)	14	18
3	–	+	–	–	13	C (psi)	60	80
4	+	+	–	–	16	D (°C)	200	250
5	–	–	+	–	17			
6	+	–	+	–	15			
7	–	+	+	–	20			
8	+	+	+	–	15			
9	–	–	–	+	10			
10	+	–	–	+	25			
11	–	+	–	+	13			
12	+	+	–	+	24			
13	–	–	+	+	19			
14	+	–	+	+	21			
15	–	+	+	+	17			
16	+	+	+	+	23			

(a) Estimate the factor effects. Based on a normal probability plot of the effect estimates, identify a model for the data from this experiment.
(b) Conduct an ANOVA based on the model identified in part (a). What are your conclusions?
(c) Analyze the residuals and comment on model adequacy.
(d) Find a regression model to predict yield in terms of the actual factor levels.
(e) Can this design be projected into a 2^3 design with two replicates? If so, sketch the design and show the average and range of the two yield values at each cube corner. Discuss the practical value of this plot.

(a)

```
 Term          Effect      Coef
Constant                 17.375
A               4.500     2.250
B               0.500     0.250
C               2.000     1.000
D               3.250     1.625
```

```
A*B          -0.750    -0.375
A*C          -4.250    -2.125
A*D           4.000     2.000
B*C           0.250     0.125
B*D           0.000     0.000
C*D           0.000     0.000
A*B*C         1.000     0.500
A*B*D         0.750     0.375
A*C*D        -0.250    -0.125
B*C*D        -0.750    -0.375
A*B*C*D       1.000     0.500
```



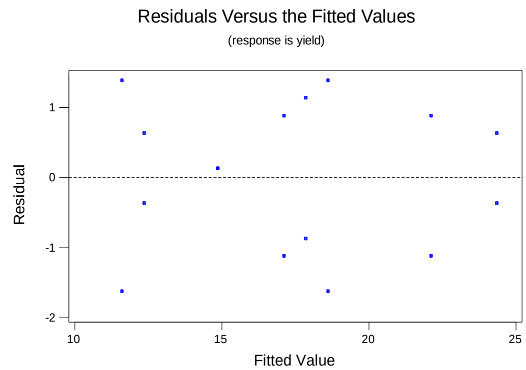
Factors A and D and interactions AC and AD are significant. For a hierarchical model, Factor C should also be included.

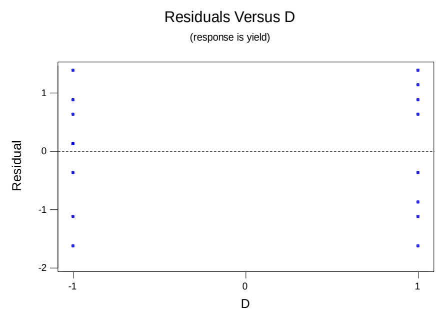
(b)

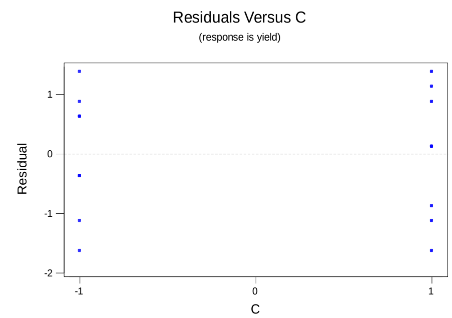
```
Analysis of Variance for yield, using Adjusted SS for Tests
Source      DF     Seq SS     Adj SS     Adj MS       F      P
A            1     81.000     81.000     81.000   49.85  0.000
C            1     16.000     16.000     16.000    9.85  0.011
D            1     42.250     42.250     42.250   26.00  0.000
A*C          1     72.250     72.250     72.250   44.46  0.000
A*D          1     64.000     64.000     64.000   39.38  0.000
Error       10     16.250     16.250      1.625
Total       15    291.750
```

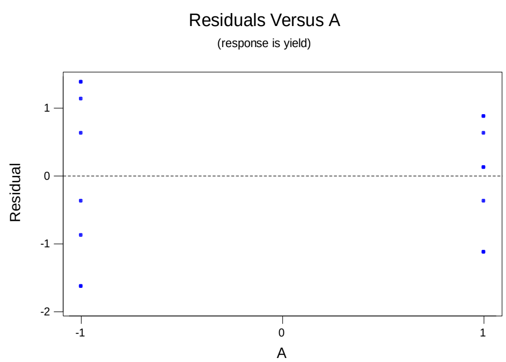
All of the factors and interactions in this table are significant at $\alpha = 0.05$.

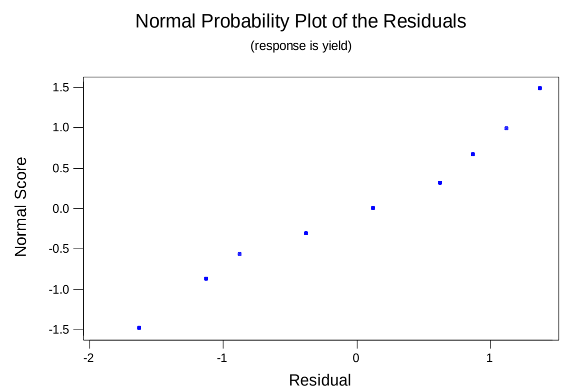
(c) The analysis of the residuals shows that the assumptions of normality and constant variance are reasonable.







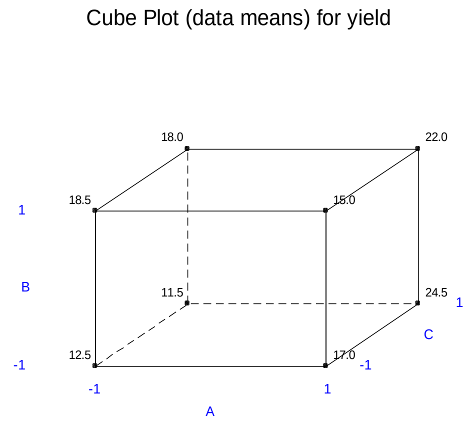




(d) The regression equation is

$$\hat{y} = 17.4 + 2.25x_1 + 1.0x_3 + 1.62x_4 - 2.12x_1x_3 + 2.00x_1x_4$$

(e) Yes, this design can be projected into a 2^3 design with 2 replicates by removing factor B.



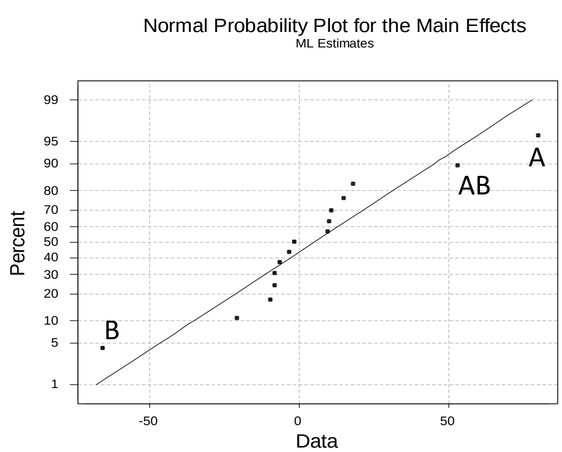
The cube plot shows the means at the high and low of each level. It can also be used to identify the interactions.

14-23 An experiment has run a single replicate of a 2^4 design and calculated the following factor effects:

A = 80.25	AB = 53.25	ABC = –2.95
B = –65.50	AC = 11.00	ABD = –8.00
C = –9.25	AD = 9.75	ACD = 10.25
D = –20.50	BC = 18.36	BCD = –7.95
	BD = 15.10	$ABCD$ = –6.25
	CD = –1.25	

(a) Construct a normal probability plot of the effects.
(b) Identify a tentative model, based on the plot of effects in part (a).
(c) Estimate the regression coefficients in this model, assuming that $\overline{y}$ = 400.

(a)



(b) Based on the normal probability plot of the effects, factors A, B and AB are significant.

(c) The estimated model is: $\hat{y} = 400 + 40.124x_1 - 32.75x_2 + 26.625x_1x_2$

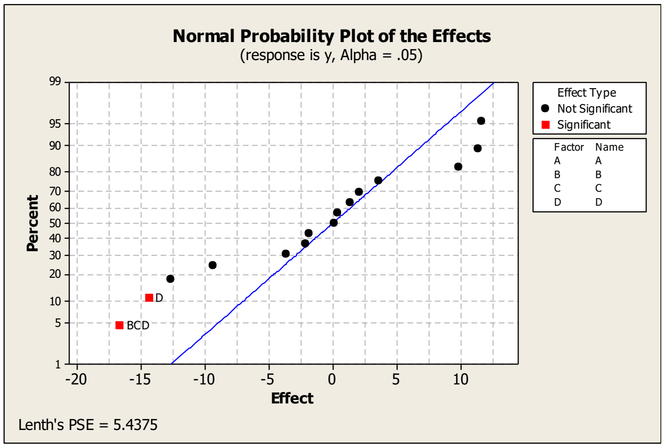
14-24 A two-level factorial experiment in four factors was conducted by Chrysler and described in the article "Sheet Molded Compound Process Improvement" by P. I. Hsieh and D. E. Goodwin (*Fourth Symposium on Taguchi Methods*, American Supplier Institute, Dearborn, MI, 1986, pp. 13–21). The purpose was to reduce the number of defects in the finish of sheet-molded grill opening panels. A portion of the experimental design, and the resulting number of defects, y_i observed on each run is shown in the following table. This is a single replicate of the 2^4 design.

(a) Estimate the factor effects and use a normal probability plot to tentatively identify the important factors.

(b) Fit an appropriate model using the factors identified in part (a).
(c) Plot the residuals from this model versus the predicted number of defects. Also prepare a normal probability plot of the residuals. Comment on the adequacy of these plots.
(d) The following table also shows the square root of the number of defects. Repeat parts (a) and (c) of the analysis using the square root of the number of defects as the response. Does this change the conclusions?

	Grill Defects Experiment					
Run	A	B	C	D	y	$\sqrt{y}$
1	–	–	–	–	56	7.48
2	+	–	–	–	17	4.12
3	–	+	–	–	2	1.41
4	+	+	–	–	4	2.00
5	–	–	+	–	3	1.73
6	+	–	+	–	4	2.00
7	–	+	+	–	50	7.07
8	+	+	+	–	2	1.41
9	–	–	–	+	1	1.00
10	+	–	–	+	0	0.00
11	–	+	–	+	3	1.73
12	+	+	–	+	12	3.46
13	–	–	+	+	3	1.73
14	+	–	+	+	4	2.00
15	–	+	+	+	0	0.00
16	+	+	+	+	0	0.00

(a)
```
Estimated Effects and Coefficients for y (coded units)

Term          Effect    Coef
Constant               10.063
A             -9.375   -4.688
B             -1.875   -0.937
C             -3.625   -1.813
D            -14.375   -7.188
A*B            0.125    0.063
A*C           -2.125   -1.062
A*D           11.625    5.812
B*C           11.375    5.688
B*D            3.625    1.813
C*D            1.375    0.688
A*B*C        -12.625   -6.313
A*B*D          2.125    1.062
A*C*D          0.375    0.188
B*C*D        -16.625   -8.313
A*B*C*D        9.875    4.938
```

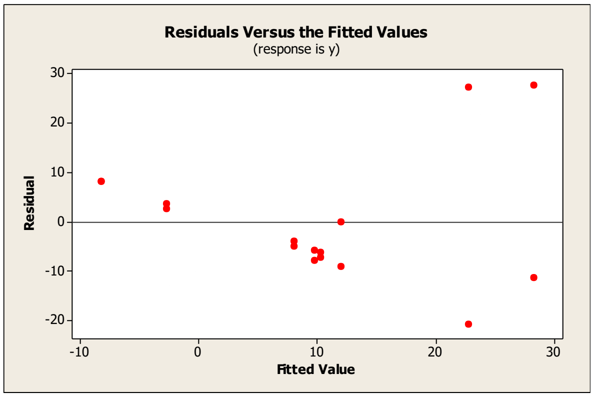


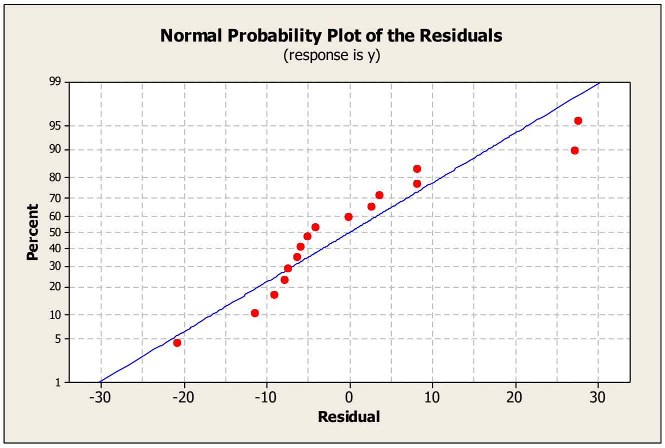
Effects D and BCD are significant effects.

(b) The model based on result from (a) is $\hat{y} = 10.063 - 7.188x_4 - 8.313x_2x_3x_4$

The hierarchical model is $\hat{y} = 10.063 - 0.937x_2 - 1.812x_3 - 7.188x_4 - 8.313x_2x_3x_4$

(c) The residual versus predicted (fitted) value shows that the model is inadequate. There is also a problem with the normality assumption as shown in normal probability plot.

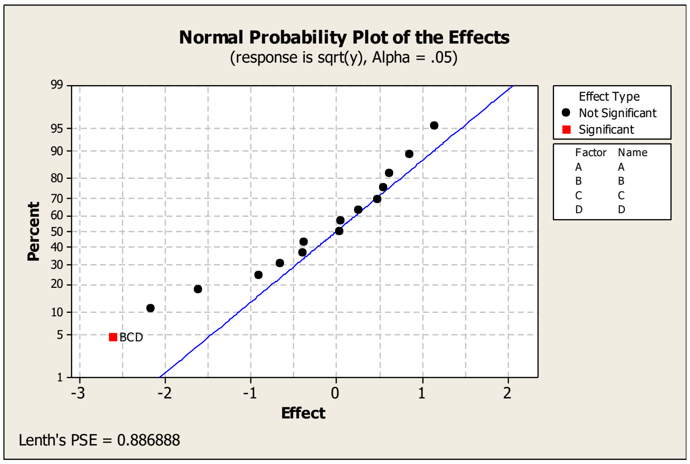




(d)

```
Estimated Effects and Coefficients for sqrt(y) (coded units)
```

Term	Effect	Coef
Constant		2.323
A	-0.895	-0.448
B	-0.372	-0.186
C	-0.658	-0.329
D	-2.164	-1.082
A*B	0.061	0.030
A*C	-0.385	-0.192
A*D	1.145	0.573
B*C	0.627	0.314
B*D	0.488	0.244
C*D	0.042	0.021
A*B*C	-1.609	-0.804
A*B*D	0.555	0.278
A*C*D	0.269	0.134
B*C*D	-2.609	-1.305
A*B*C*D	0.859	0.429



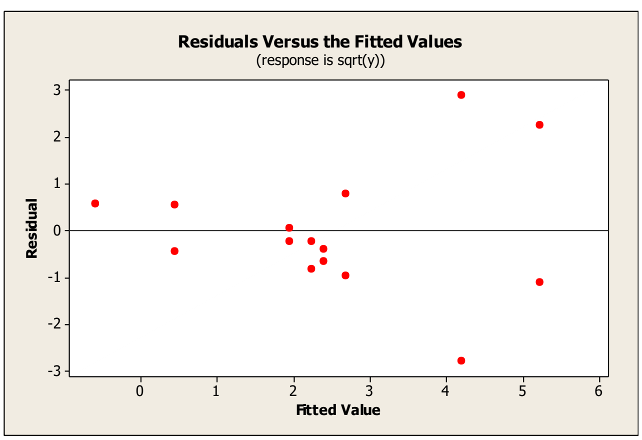
From the normal probability plot, only BCD is a significant effect.

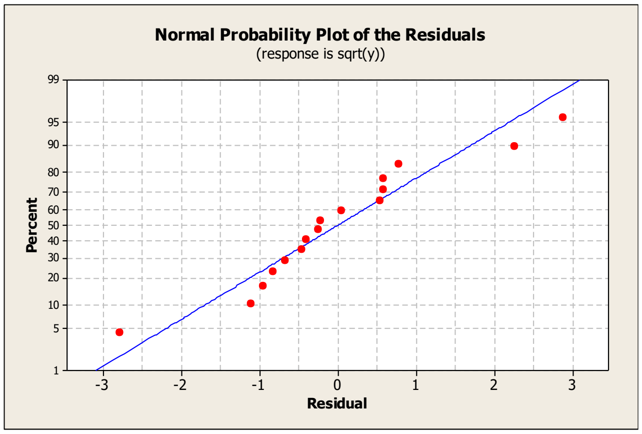
(b) The model based on result from (a) is $\hat{y} = 2.323 - 1.305 x_2 x_3 x_4$

A model with main effects, but without the two-factor interactions, is

$$\hat{y} = 2.323 - 0.186 x_2 - 0.329 x_3 - 1.082 x_4 - 1.305 x_2 x_3 x_4$$

(c) The plots look very similar to the residual plots from the untransformed data.





14-25 Consider a 2^2 factorial experiment with four center points. The data are (1) = 21, *a* = 125, *b* = 154, *ab* = 352, and the responses at the center point are 92, 130, 98, 152. Compute an ANOVA with the sum of squares for curvature and conduct an *F*-test for curvature. Use α = 0.05.

Only main effects are significant. Interaction effects and curvature are not significant at α = 0.05.

Analysis of Variance for Response (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	2	55201	55201	27600.5	34.85	0.008
2-Way Interactions	1	2209	2209	2209.0	2.79	0.193
Curvature	1	4050	4050	4050.0	5.11	0.109
Residual Error	3	2376	2376	792.0		
Pure Error	3	2376	2376	792.0		
Total	7	63836				

For the curvature, because $F_0 = 5.11 < F_{0.05,1,3} = 10.13$ so there is no evidence that curvature is significant at α = 0.05.

14-26 Consider the experiment in Exercise 14-16. Suppose that a center point with five replicates is added to the factorial runs and the responses are 2800, 5600, 4500, 5400, 3600. Compute an ANOVA with the sum of squares for curvature and conduct an *F*-test for curvature. Use α = 0.05.

Only curvature is significant at α = 0.05.

Analysis of Variance for Strength (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	120635081	120635081	24127016	17.09	0.008
2-Way Interactions	10	12316561	12316561	1231656	0.87	0.610
3-Way Interactions	10	724711	724711	72471	0.05	1.000
4-Way Interactions	5	188406	188406	37681	0.03	0.999
5-Way Interactions	1	31626	31626	31626	0.02	0.888
Curvature	1	9630256	9630256	9630256	6.82	0.059
Residual Error	4	5648000	5648000	1412000		
Pure Error	4	5648000	5648000	1412000		
Total	36	149174640				

For the curvature, because $F_0 = 6.82 < F_{0.05,1,4} = 7.71$ so there is no evidence that curvature is significant at α = 0.05.

14-27 Consider the experiment in Exercise 14-19. Suppose that a center point with five replicates is added to the factorial runs and the responses are 45, 40, 41, 47, and 43.

(a) Estimate the experimental error using the center points. Compare this to the estimate obtained originally in Exercise 14-19 by pooling apparently nonsignificant effects.
(b) Test for curvature with $\alpha = 0.05$.

(a) Original: From original analysis terms A, B, C, and AB are significant

Factorial Fit: y versus A, B, C

```
Estimated Effects and Coefficients for y (coded units)

Term      Effect    Coef  SE Coef       T      P
Constant          30.531   0.3021  101.07  0.000
A         11.813   5.906   0.3021   19.55  0.000
B         33.937  16.969   0.3021   56.17  0.000
C          9.688   4.844   0.3021   16.03  0.000
A*B        7.938   3.969   0.3021   13.14  0.000


S = 1.70884   R-Sq = 99.32%   R-Sq(adj) = 99.22%

Analysis of Variance for y (coded units)

Source              DF   Seq SS   Adj SS   Adj MS        F      P
Main Effects         3  11081.1  11081.1  3693.70  1264.90  0.000
2-Way Interactions   1    504.0    504.0   504.03   172.61  0.000
Residual Error      27     78.8     78.8     2.92
  Lack of Fit        3      3.1      3.1     1.03     0.33  0.806
  Pure Error        24     75.8     75.8     3.16
Total               31  11664.0
```

With 5 center points:
The standard deviation of the 5 center points is 2.86 and this is an estimate of experimental error. This is similar to the estimate s =1.71 from the original ANOVA.

If the original ANOVA is combined with these center points the ANOVA below is generated. The center points contribute an additional 4 degrees of freedom to the residual error.

Factorial Fit: y versus A, B, C

```
Estimated Effects and Coefficients for y (coded units)

Term      Effect    Coef  SE Coef      T      P
Constant          30.531   0.3355  91.01  0.000
A         11.813   5.906   0.3355  17.61  0.000
B         33.937  16.969   0.3355  50.58  0.000
C          9.688   4.844   0.3355  14.44  0.000
A*B        7.938   3.969   0.3355  11.83  0.000
Ct Pt             12.669   0.9126  13.88  0.000

S = 1.89774   R-Sq = 99.10%   R-Sq(adj) = 98.95%

Analysis of Variance for y (coded units)

Source              DF   Seq SS   Adj SS   Adj MS        F      P
Main Effects         3  11081.1  11081.1  3693.70  1025.63  0.000
2-Way Interactions   1    504.0    504.0   504.03   139.95  0.000
  Curvature          1    694.0    694.0   694.04   192.71  0.000
Residual Error      31    111.6    111.6     3.60
```

```
  Lack of Fit          3      3.1      3.1     1.03     0.27  0.849
  Pure Error          28    108.6    108.6     3.88
Total                 36  12390.8
```

(b) From the last ANOVA above, the *P*-value for curvature is near zero. There is strong evidence of curvature from this data.

14-28 An article in *Talanta* (2005, Vol. 65, pp. 895–899) presented a 2^3 factorial design to find lead level by using flame atomic absorption spectrometry (FAAS). The data are in the following table.

	Factors			Lead Recovery (%)	
Run	**ST**	**pH**	**RC**	**R1**	**R2**
1	–	–	–	39.8	42.1
2	+	–	–	51.3	48
3	–	+	–	57.9	58.1
4	+	+	–	78.9	85.9
5	–	–	+	78.9	84.2
6	+	–	+	84.2	84.2
7	–	+	+	94.4	90.9
8	+	+	+	94.7	105.3

The factors and levels are in the following table.

Factor	Low (–)	High (+)
Reagent concentration (RC) (mol 1^{-1})	5×10^{-6}	5×10^{-5}
pH	6.0	8.0
Shaking time (ST) (min)	10	30

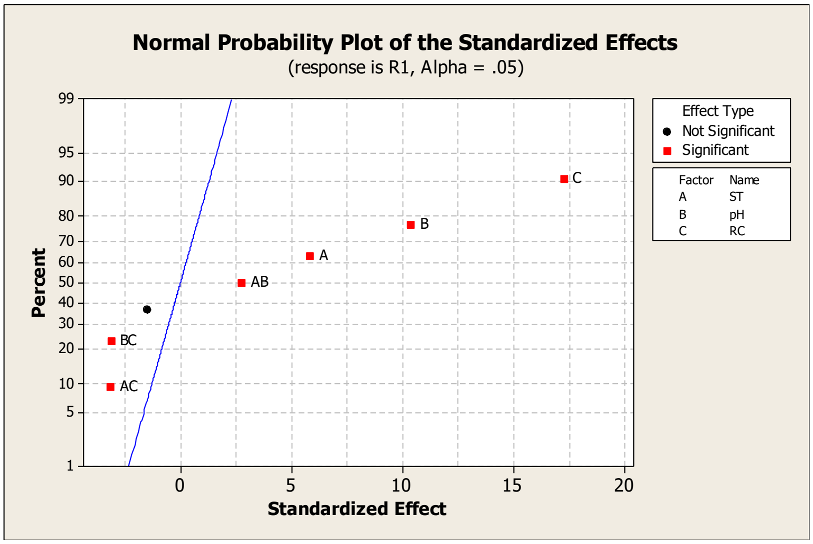
(a) Construct a normal probability plot of the effect estimates. Which effects appear to be large?
(b) Conduct an analysis of variance to confirm your findings for part (a).
(c) Analyze the residuals from this experiment. Are there any problems with model adequacy?

(a) The plot below shows standardized effects. A standardized effect equals the effect divided by its standard error estimate. Because an effect equals two times its coefficient, the standard error estimate for an effect equals two times the standard error estimate for its coefficient and these are provided in the table. The standardized effect for ST = 10.775/[2(0.9226)] = 17.2.

The effects of all main factors (ST, pH, and RC) and two-factor interaction terms (ST*pH, ST*RC, and pH*RC) are large.

Factorial Fit: R1 versus ST, pH, RC

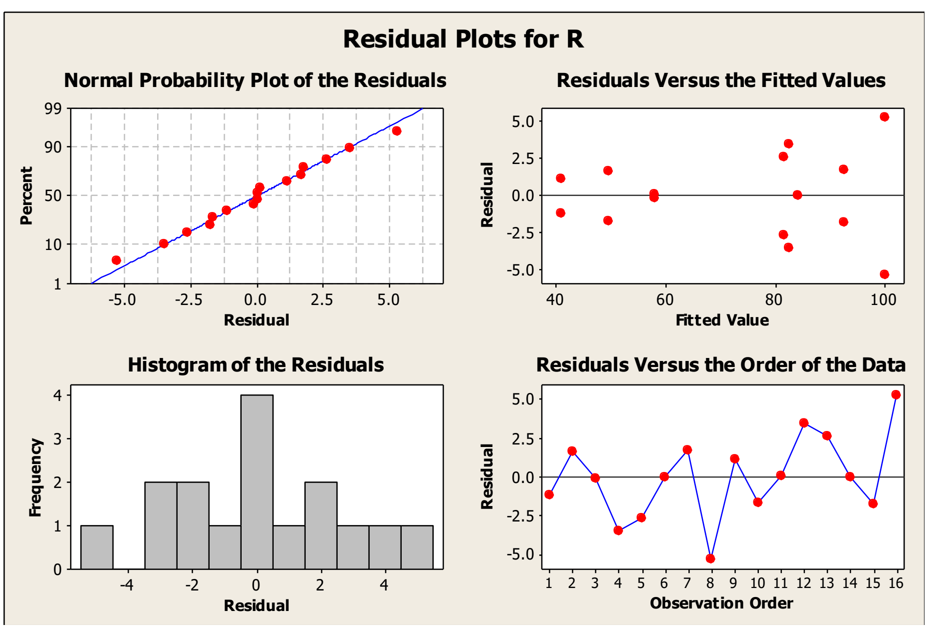
```
Estimated Effects and Coefficients for R1 (coded units)

Term        Effect    Coef  SE Coef      T      P
Constant            73.675   0.9226  79.85  0.000
ST          10.775   5.387   0.9226   5.84  0.000
pH          19.175   9.587   0.9226  10.39  0.000
RC          31.850  15.925   0.9226  17.26  0.000
ST*pH        5.100   2.550   0.9226   2.76  0.025
ST*RC       -5.775  -2.888   0.9226  -3.13  0.014
pH*RC       -5.725  -2.863   0.9226  -3.10  0.015
ST*pH*RC    -2.750  -1.375   0.9226  -1.49  0.174

S = 3.69053   R-Sq = 98.32%   R-Sq(adj) = 96.86%
```



(b) Computer output below combines the sum of squares and the degrees of freedom for the main effects, the two-factor effects, and the three factor effects. An F statistic for each individual effect can be obtained from the square of the t statistic in the previous table. That is, the F statistic for ST = $79.85^2 = 6376.02$. However, because the *P*-value for each F test is the same as the P-value for the t test, the test for each effect is already provided with the t statistics.

```
Analysis of Variance for R1 (coded units)

Source               DF   Seq SS   Adj SS   Adj MS       F      P
Main Effects          3  5992.81  5992.81  1997.60  146.67  0.000
2-Way Interactions    3   368.55   368.55   122.85    9.02  0.006
3-Way Interactions    1    30.25    30.25    30.25    2.22  0.174
Residual Error        8   108.96   108.96    13.62
  Pure Error          8   108.96   108.96    13.62
Total                15  6500.57
```

(c) The normality assumption is reasonable. The plot of residuals versus the predicted values indicates some greater variability for larger fitted values so that some departure from assumptions is indicated. The actual time order of the observations was not provided so the plot versus observation order is not relevant.



14-29 An experiment to study the effect of machining factors on ceramic strength was described at www.itl.nist.gov/div898/handbook/. Five factors were considered at two levels each: A = Table Speed, B = Down Feed Rate, C = Wheel Grit, D = Direction, E = Batch. The response is the average of the ceramic strength over 15 repetitions. The following data are from a single replicate of a 2^5 factorial design.

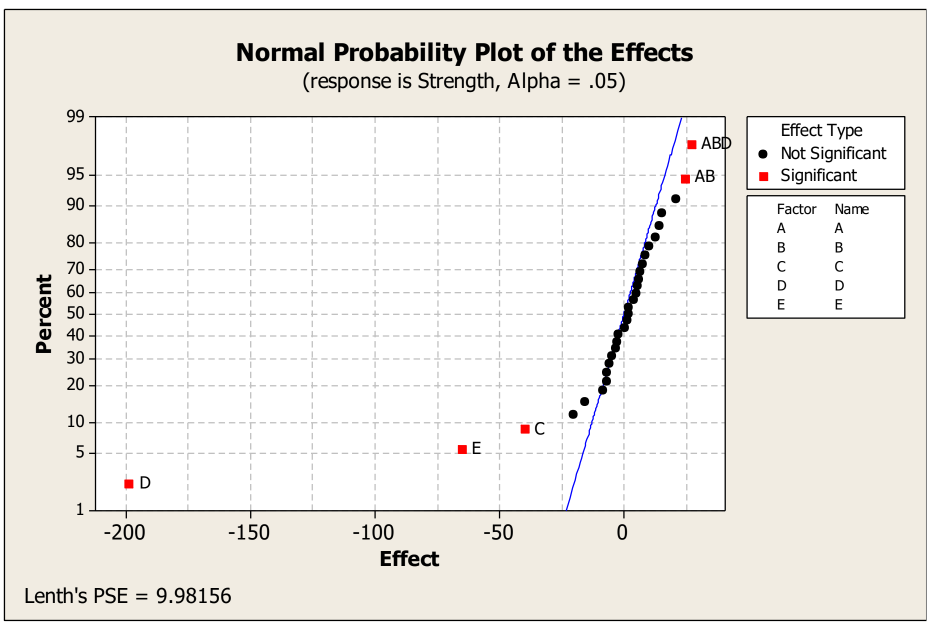
A	B	C	D	E	Strength
–1	–1	–1	–1	–1	680.45
1	–1	–1	–1	–1	722.48
–1	1	–1	–1	–1	702.14
1	1	–1	–1	–1	666.93
–1	–1	1	–1	–1	703.67
1	–1	1	–1	–1	642.14
–1	1	1	–1	–1	692.98
1	1	1	–1	–1	669.26
–1	–1	–1	1	–1	491.58
1	–1	–1	1	–1	475.52
–1	1	–1	1	–1	478.76
1	1	–1	1	–1	568.23
–1	–1	1	1	–1	444.72
1	–1	1	1	–1	410.37
–1	1	1	1	–1	428.51
1	1	1	1	–1	491.47
–1	–1	–1	–1	1	607.34
1	–1	–1	–1	1	620.8
–1	1	–1	–1	1	610.55
1	1	–1	–1	1	638.04
–1	–1	1	–1	1	585.19
1	–1	1	–1	1	586.17
–1	1	1	–1	1	601.67
1	1	1	–1	1	608.31
–1	–1	–1	1	1	442.9
1	–1	–1	1	1	434.41
–1	1	–1	1	1	417.66
1	1	–1	1	1	510.84
–1	–1	1	1	1	392.11
1	–1	1	1	1	343.22
–1	1	1	1	1	385.52
1	1	1	1	1	446.73

(a) Estimate the factor effects and use a normal probability plot of the effects. Identify which effects appear to be large.
(b) Fit an appropriate model using the factors identified in part (a).
(c) Prepare a normal probability plot of the residuals. Also, plot the residuals versus the predicted ceramic strength. Comment on the adequacy of these plots.
(d) Identify and interpret any significant interactions.
(e) What are your recommendations regarding process operating conditions?

(a)
Factorial Fit: Strength versus A, B, C, D, E

```
Estimated Effects and Coefficients for Strength (coded units)
Term         Effect    Coef
Constant             546.90
A             10.57    5.29
```

B	20.91	10.45
C	-39.79	-19.89
D	-198.47	-99.24
E	-64.86	-32.43
A*B	24.68	12.34
A*C	-15.16	-7.58
A*D	14.31	7.15
A*E	7.62	3.81
B*C	6.20	3.10
B*D	15.70	7.85
B*E	4.99	2.49
C*D	-19.87	-9.93
C*E	-1.92	-0.96
D*E	12.89	6.44
A*B*C	6.68	3.34
A*B*D	27.15	13.57
A*C*D	0.51	0.26
A*B*E	4.25	2.13
A*C*E	1.95	0.97
A*D*E	-8.25	-4.13
B*C*D	-2.36	-1.18
B*C*E	1.79	0.89
B*D*E	-4.57	-2.29
C*D*E	2.01	1.01
A*B*C*D	-6.65	-3.33
A*B*C*E	-6.67	-3.34
A*B*D*E	-3.14	-1.57
A*C*D*E	-5.39	-2.70
B*C*D*E	5.80	2.90
A*B*C*D*E	8.75	4.38



The effects for factors C, D, E are large. The AB and ABD effects are next largest in magnitude and might be considered significant.

(b)

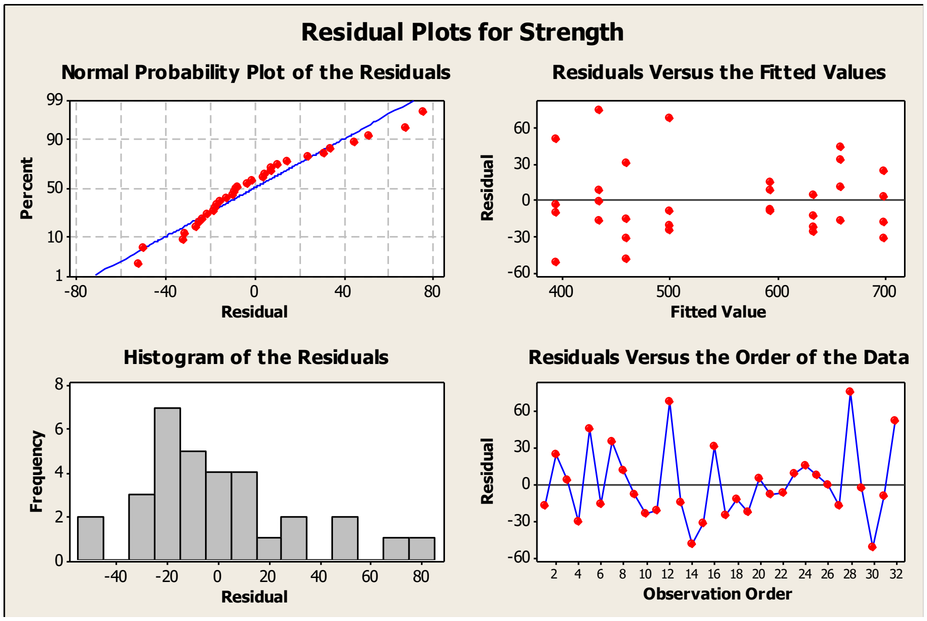
Factorial Fit: Strength versus C, D, E

```
Estimated Effects and Coefficients for Strength (coded units)

Term        Effect    Coef  SE Coef       T      P
Constant            546.90    5.723   95.56  0.000
C           -39.79  -19.89    5.723   -3.48  0.002
D          -198.47  -99.24    5.723  -17.34  0.000
E           -64.86  -32.43    5.723   -5.67  0.000

S = 32.3760   R-Sq = 92.49%   R-Sq(adj) = 91.69%

Analysis of Variance for Strength (coded units)

Source           DF  Seq SS  Adj SS  Adj MS       F      P
Main Effects      3  361451  361451  120484  114.94  0.000
Residual Error   28   29350   29350    1048
  Lack of Fit     4    4549    4549    1137    1.10  0.379
  Pure Error     24   24801   24801    1033
Total            31  390800
```

The average of the ceramic strength is given by

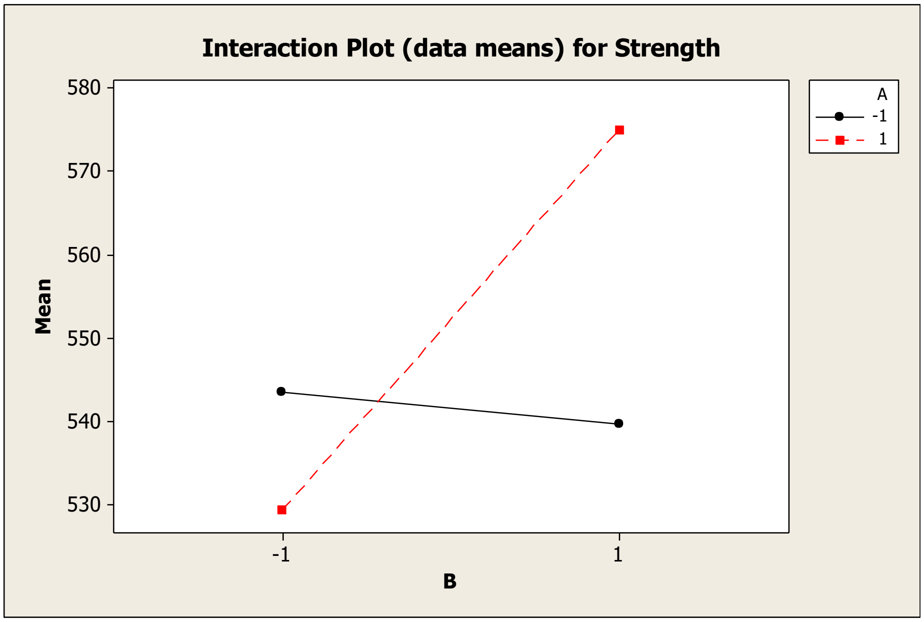
$$\hat{y} = 546.90 - 19.89C - 99.24D - 32.43E$$

(c)



The normality assumption is reasonable. The plot of residuals versus the predicted values indicates some greater variability for smaller fitted values, but a strong departure from assumptions is not indicated. The actual time order of the observations was not provided so the plot versus observation order is not relevant.

(d) The two interaction terms in this model AB and ABD are large, but not considered significant. We illustrate plots for these effects below. The plot of the AB interaction shows that the effect of changing factor B at low level of A is small, but increasing factor B at high level of A has a greater effect on the average of the ceramic strength (but still not significant).



(e) Use the model with only effects C, D, and E and assume that the objective of the process is to maximize the average of the ceramic strength. It can be seen from the equation $\hat{y} = 546.90 - 19.89C - 99.24D - 32.43E$ that the settings for the factors should be C = -1, D = -1, and E = -1. Factors A and B are not important either as main effects or interactions, so these may be set at any convenient levels.

14-30 Consider the following computer output for a 2^3 factorial experiment.

(a) How many replicates were used in the experiment?
(b) Use the appropriate equation to calculate the standard error of a coefficient.
(c) Calculate the entries marked with "?" in the output.

Estimated Effects and Coefficients

Term	Effect	Coef	SE Coef	t	P
Constant		583.57	?	115.40	0.000
A	5.40	?	?	?	?
B	20.94	10.47	?	2.07	0.072
C	–41.73	–20.86	?	–4.13	0.003
A*B	26.93	13.46	?	2.66	0.029
A*C	–20.41	–10.20	?	–2.02	0.078
B*C	3.91	1.96	?	0.39	0.709
A*B*C	–12.07	–6.04	?	–1.19	0.267

S = 20.2279

Analysis of Variance

Source	DF	SS	MS	F	P
A	?	?	?	?	?
B	?	1753.4	1753.40	4.29	0.072
C	?	6965.0	6964.95	17.02	0.003
A*B	?	2900.8	2900.76	7.09	0.029
A*C	?	1665.9	1665.93	4.07	0.078
B*C	?	61.3	61.25	0.15	0.709
A*B*C	?	583.1	583.06	1.42	0.267
Residual error	?	3273.4	409.17		
Total	15	17319.5			

(a) Because the degrees of freedom total = 15, there are 16 trials. There are three factors: A, B, and C with two levels each. Therefore, there are 2 replicates used in this experiment.

(b) SE Coef = $\hat{\sigma}\sqrt{\dfrac{1}{n2^k}} = \sqrt{23664.2}\sqrt{\dfrac{1}{2(2^3)}} = 38.46$

(c) Coef of B = $\dfrac{Effect}{2} = \dfrac{15.92}{2} = 7.96$

T statistics for A*B = $\dfrac{\hat{\beta}}{standard\ \ error\ \hat{\beta}} = \dfrac{10.21}{38.46} = 0.27$

SE Coef of A*B*C = $\hat{\sigma}\sqrt{\dfrac{1}{n2^k}} = \sqrt{23664.2}\sqrt{\dfrac{1}{2(2^3)}} = 38.46$

F for Main Effects = (Adj MS of Main effects)/ (Adj MS of Error) = 2261.8/23664.2 = 0.09558

Seq SS of 2-way interactions = Adj SS of 2-way interactions = 2918
Seq SS of 3-way interactions = Adj SS of 3-way interactions = 713

14-31 Consider the following computer output for one replicate of a 2^4 factorial experiment.

(a) What effects are used to estimate error?
(b) Calculate the entries marked with "?" in the output.

Estimated Effects and Coefficients

Term	Effect	Coef	SE Coef	t	P
Constant		35.250	?	39.26	0.000
A	2.250	?	?	?	?
B	24.750	12.375	?	13.78	0.000
C	1.000	0.500	?	0.56	0.602
D	10.750	5.375	?	5.99	0.002
A*B	-10.500	-5.250	?	-5.85	0.002
A*C	4.250	2.125	?	2.37	0.064
A*D	-5.000	-2.500	?	-2.78	0.039
B*C	5.250	2.625	?	2.92	0.033
B*D	4.000	2.000	?	2.23	0.076
C*D	-0.750	-0.375	?	-0.42	0.694

S = 3.59166

Analysis of Variance

Source	DF	SS	MS	F	P
A	?	?	?	?	0.266
B	1	2450.25	2450.25	189.94	0.000
C	1	4.00	4.00	0.31	0.602
D	1	462.25	462.25	35.83	0.002
AB	1	441.00	441.00	34.19	0.002
AC	1	72.25	72.25	5.60	0.064
AD	1	100.00	100.00	7.75	0.039
BC	1	110.25	110.25	8.55	0.033
BD	1	64.00	64.00	4.96	0.076
CD	1	2.25	2.25	0.17	0.694
Residual Error	?	64.50	?		
Total	?	3791.00			

(a) Interaction effects with more than two factors are used to estimate error.

(b) Because this is a single replicate of a 2^4 experiment, there are 16 tests total. Therefore, df(Total) = 15. Because there are 10 effects in the model df(Error) = 15 – 10 = 5 and MS(Error) = 64.50/5 = 12.9.

Therefore, the standard error for a coefficient = $0.5[12.9(1/8 + 1/8)]^{1/2} = 0.90$

The *t* test for A is 1.125/0.90 = 1.25. From the *t* distribution with 5 degrees of freedom, this corresponds to a two-sided probability of 0.267.

Because the set of sums of squares needs to add to SS(Total), $SS_A = 20.25$, with 1 degree of freedom. Therefore, $MS_A = 20.25$ and F = 20.25/12.9 = 1.57

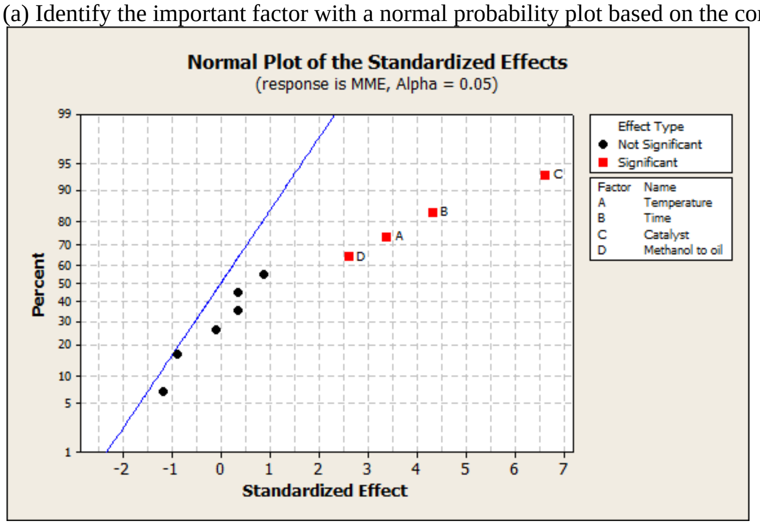
```
Estimated Effects and Coefficients
Term          Effect     Coef  SE Coef  t       P
Constant               35.250    0.90   39.26  0.000
A              2.250    1.125    0.90    1.25  0.267
B             24.750   12.375    0.90   13.78  0.000
C              1.000    0.500    0.90    0.56  0.602
D             10.750    5.375    0.90    5.99  0.002
A*B          -10.500   -5.250    0.90   -5.85  0.002
A*C            4.250    2.125    0.90    2.37  0.064
A*D           -5.000   -2.500    0.90   -2.78  0.039
B*C            5.250    2.625    0.90    2.92  0.033
B*D            4.000    2.000    0.90    2.23  0.076
C*D           -0.750   -0.375    0.90   -0.42  0.694

S = 3.59166

Analysis of Variance
Source                  DF    SS          MS         F      P
  A                     1     20.25       20.25      1.57   0.266
  B                     1   2450.25     2450.25    189.94   0.000
  C                     1      4.00        4.00      0.31   0.602
  D                     1    462.25      462.25     35.83   0.002
  A*B                   1    441.00      441.00     34.19   0.002
  A*C                   1     72.25       72.25      5.60   0.064
  A*D                   1    100.00      100.00      7.75   0.039
  B*C                   1    110.25      110.25      8.55   0.033
  B*D                   1     64.00       64.00      4.96   0.076
  C*D                   1      2.25        2.25      0.17   0.694
Residual Error          5     64.50       12.9
Total                   15  3791.00
```

14-32 An article in *Bioresource Technology* ("Influence of Vegetable Oils Fatty-Acid Composition on Biodiesel Optimization," (2011, Vol. 102(2), pp. 1059–1065)] described an experiment to analyze the influence of the fatty-acid composition on biodiesel. Factors were the concentration of catalyst, amount of methanol, reaction temperature and time, and the design included three center points. Maize oil methyl ester (MME) was recorded as the response. Data follow.

Run	Temperature (°C)	Time (min)	Catalyst (wt.%)	Methanol to oil molar ratio	MME (wt.%)
1	45	40	0.8	5.4	88.30
2	25	40	1.2	5.4	90.50
3	45	10	0.8	4.2	77.96
4	25	10	1.2	5.4	85.59
5	45	40	1.2	5.4	97.14
6	45	10	1.2	4.2	90.64
7	45	40	1.2	4.2	89.86
8	25	40	0.8	4.2	82.35
9	25	10	0.8	5.4	80.31
10	25	40	0.8	5.4	85.51
11	25	10	0.8	4.2	76.21
12	45	40	0.8	4.2	86.86
13	25	10	1.2	4.2	86.35
14	45	10	0.8	5.4	84.58
15	25	40	1.2	4.2	89.37
16	45	10	1.2	5.4	90.51
17	35	25	1	4.8	91.40
18	35	25	1	4.8	91.96
19	35	25	1	4.8	91.07

(a) Identify the important effects from a normal probability plot.
(b) Compare the results in the previous part with results that use an error term based on the center points.
(c) Test for curvature.
(d) Analyze the residuals from the model.

(a) Identify the important factor with a normal probability plot based on the corner points.



(b) Compare the results in the previous part with results that use an error term based on the center points.

```
Estimated Effects and Coefficients for MME (coded units)

Term                          Effect     Coef  SE Coef       T      P
Constant                              86.3775   0.1125  767.93  0.000
Temperature                   3.7075   1.8537   0.1125   16.48  0.004
Time                          4.7175   2.3587   0.1125   20.97  0.002
Catalyst                      7.2350   3.6175   0.1125   32.16  0.001
Ratio                         2.8550   1.4275   0.1125   12.69  0.006
Temperature*Time             -0.1000  -0.0500   0.1125   -0.44  0.700
Temperature*Catalyst          0.3775   0.1888   0.1125    1.68  0.235
```

Temperature*Ratio	0.9475	0.4737	0.1125	4.21	0.052
Time*Catalyst	-1.2725	-0.6363	0.1125	-5.66	0.030
Time*Ratio	0.3975	0.1987	0.1125	1.77	0.219
Catalyst*Ratio	-0.9750	-0.4875	0.1125	-4.33	0.049
Temperature*Time*Catalyst	-0.4200	-0.2100	0.1125	-1.87	0.203
Temperature*Time*Ratio	0.1600	0.0800	0.1125	0.71	0.551
Temperature*Catalyst*Ratio	0.7475	0.3738	0.1125	3.32	0.080
Time*Catalyst*Ratio	1.9275	0.9637	0.1125	8.57	0.013
Temperature*Time*Catalyst*Ratio	1.2200	0.6100	0.1125	5.42	0.032
Ct Pt		5.0992	0.2831	18.01	0.003

S = 0.449926   PRESS = *
R-Sq = 99.92%   R-Sq(pred) = *%   R-Sq(adj) = 99.26%

Analysis of Variance for MME (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F
Main Effects	4	385.986	385.986	96.497	476.68
Temperature	1	54.982	54.982	54.982	271.61
Time	1	89.019	89.019	89.019	439.75
Catalyst	1	209.381	209.381	209.381	1034.32
Ratio	1	32.604	32.604	32.604	161.06
2-Way Interactions	6	15.113	15.113	2.519	12.44
Temperature*Time	1	0.040	0.040	0.040	0.20
Temperature*Catalyst	1	0.570	0.570	0.570	2.82
Temperature*Ratio	1	3.591	3.591	3.591	17.74
Time*Catalyst	1	6.477	6.477	6.477	32.00
Time*Ratio	1	0.632	0.632	0.632	3.12
Catalyst*Ratio	1	3.803	3.803	3.803	18.78
3-Way Interactions	4	17.904	17.904	4.476	22.11
Temperature*Time*Catalyst	1	0.706	0.706	0.706	3.49
Temperature*Time*Ratio	1	0.102	0.102	0.102	0.51
Temperature*Catalyst*Ratio	1	2.235	2.235	2.235	11.04
Time*Catalyst*Ratio	1	14.861	14.861	14.861	73.41
4-Way Interactions	1	5.954	5.954	5.954	29.41
Temperature*Time*Catalyst*Ratio	1	5.954	5.954	5.954	29.41
Curvature	1	65.688	65.688	65.688	324.49
Residual Error	2	0.405	0.405	0.202	
Pure Error	2	0.405	0.405	0.202	
Total	18	491.050			

Source	P
Main Effects	0.002
Temperature	0.004
Time	0.002
Catalyst	0.001
Ratio	0.006
2-Way Interactions	0.076
Temperature*Time	0.700
Temperature*Catalyst	0.235
Temperature*Ratio	0.052
Time*Catalyst	0.030
Time*Ratio	0.219
Catalyst*Ratio	0.049
3-Way Interactions	0.044
Temperature*Time*Catalyst	0.203
Temperature*Time*Ratio	0.551
Temperature*Catalyst*Ratio	0.080
Time*Catalyst*Ratio	0.013
4-Way Interactions	0.032
Temperature*Time*Catalyst*Ratio	0.032

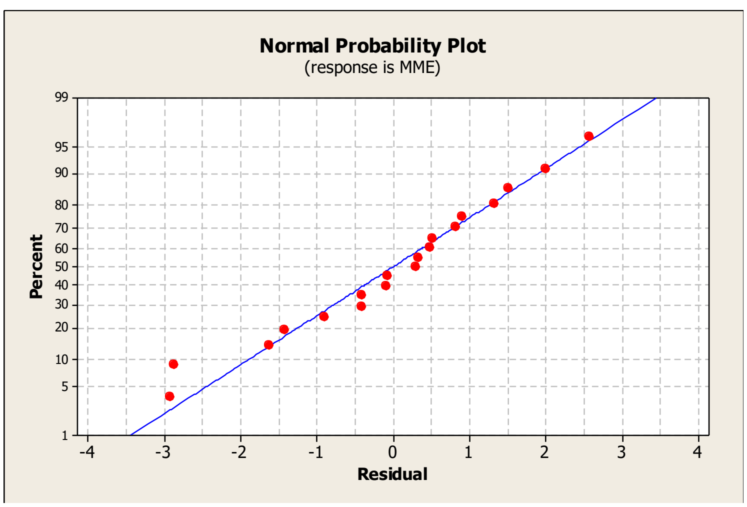
```
  Curvature                           0.003
Residual Error
  Pure Error
Total
```

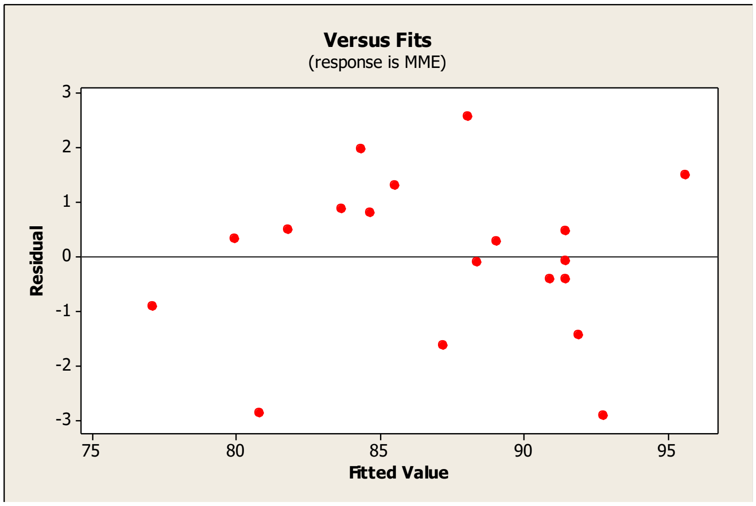
The results based on the error estimate from the center points and the normal probability plot are similar. Some interaction effects have p-values less than 0.05, but the magnitudes of these effects are much smaller than the four main effects (with much lower p values).

(c) For the curvature test, the P-value = 0.003. Therefore, significant curvature is present.

(d) A model with only the four main effects and the center point term is used to generate residuals.

```
Estimated Effects and Coefficients for MME (coded units)

Term          Effect     Coef  SE Coef       T      P
Constant               86.377   0.4351  198.53  0.000
Temperature    3.707    1.854   0.4351    4.26  0.001
Time           4.717    2.359   0.4351    5.42  0.000
Catalyst       7.235    3.617   0.4351    8.31  0.000
Ratio          2.855    1.428   0.4351    3.28  0.006
Ct Pt                   5.099   1.0950    4.66  0.000

S = 1.74036     PRESS = 83.3604
R-Sq = 91.98%   R-Sq(pred) = 83.02%   R-Sq(adj) = 88.90%

Analysis of Variance for MME (coded units)

Source          DF   Seq SS   Adj SS   Adj MS      F      P
Main Effects     4  385.986  385.986   96.497  31.86  0.000
  Temperature    1   54.982   54.982   54.982  18.15  0.001
  Time           1   89.019   89.019   89.019  29.39  0.000
  Catalyst       1  209.381  209.381  209.381  69.13  0.000
  Ratio          1   32.604   32.604   32.604  10.76  0.006
  Curvature      1   65.688   65.688   65.688  21.69  0.000
Residual Error  13   39.375   39.375    3.029
  Lack of Fit   11   38.970   38.970    3.543  17.50  0.055
  Pure Error     2    0.405    0.405    0.202
Total           18  491.050
```





There are no obvious departures from assumptions seen in these plots.

14-33 An article in *Analytica Chimica Acta* ["Design-of-Experiment Optimization of Exhaled Breath Condensate Analysis Using a Miniature Differential Mobility Spectrometer (DMS)" (2008, Vol. 628(2), pp. 155–161)] examined four parameters that affect the sensitivity and detection of the analytical instruments used to measure clinical samples. They optimized the sensor function using exhaled breath condensate (EBC) samples spiked with acetone, a known clinical biomarker in breath. The following table shows the results for a single replicate of a 2^4 factorial experiment for one of the outputs, the average amplitude of acetone peak over three repetitions.

Configuration	A	B	C	D	*y*
1	+	+	+	+	0.12
2	+	+	+	–	0.1193
3	+	+	–	+	0.1196
4	+	+	–	–	0.1192
5	+	–	+	+	0.1186
6	+	–	+	–	0.1188
7	+	–	–	+	0.1191
8	+	–	–	–	0.1186
9	–	+	+	+	0.121
10	–	+	+	–	0.1195
11	–	+	–	+	0.1196
12	–	+	–	–	0.1191
13	–	–	+	+	0.1192
14	–	–	+	–	0.1194
15	–	–	–	+	0.1188
16	–	–	–	–	0.1188

The factors and levels are shown in the following table.

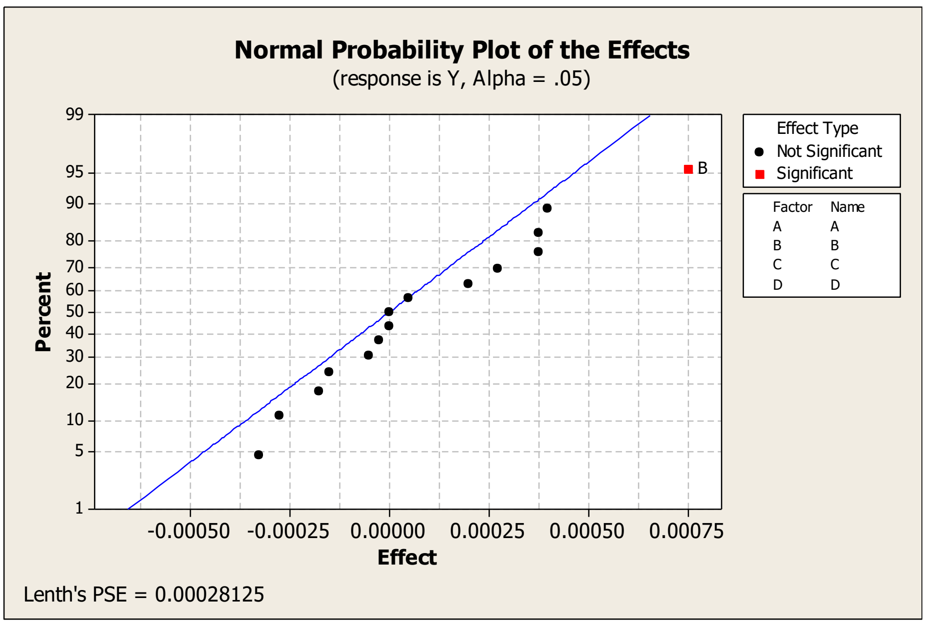
A	RF voltage of the DMS sensor (1200 or 1400 V)
B	Nitrogen carrier gas flow rate (250 or 500 mL min^{-1})
C	Solid phase microextraction (SPME) filter type (polyacrylate or PDMS–DVB)
D	GC cooling profile (cryogenic and noncryogenic)

(a) Estimate the factor effects and use a normal probability plot of the effects. Identify which effects appear to be large, and identify a model for the data from this experiment.
(b) Conduct an ANOVA based on the model identified in part (a). What are your conclusions?
(c) Analyze the residuals from this experiment. Are there any problems with model adequacy?

(d) Project the design in this problem into a 2^r design for $r < 4$ in the important factors. Sketch the design and show the average and range of yields at each run. Does this sketch aid in data representation?

(a)

```
Factorial Fit: Y versus A, B, C, D

Estimated Effects and Coefficients for Y (coded units)
Term          Effect        Coef
Constant                 0.119288
A          -0.000275   -0.000138
B           0.000750    0.000375
C           0.000375    0.000188
D           0.000400    0.000200
A*B         0.000000    0.000000
A*C        -0.000325   -0.000162
A*D        -0.000050   -0.000025
B*C         0.000200    0.000100
B*D         0.000375    0.000188
C*D         0.000050    0.000025
A*B*C       0.000000    0.000000
A*B*D      -0.000175   -0.000088
A*C*D      -0.000150   -0.000075
B*C*D       0.000275    0.000138
A*B*C*D    -0.000025   -0.000013
```



The effect of factor B is large, so this factor is included in the model.

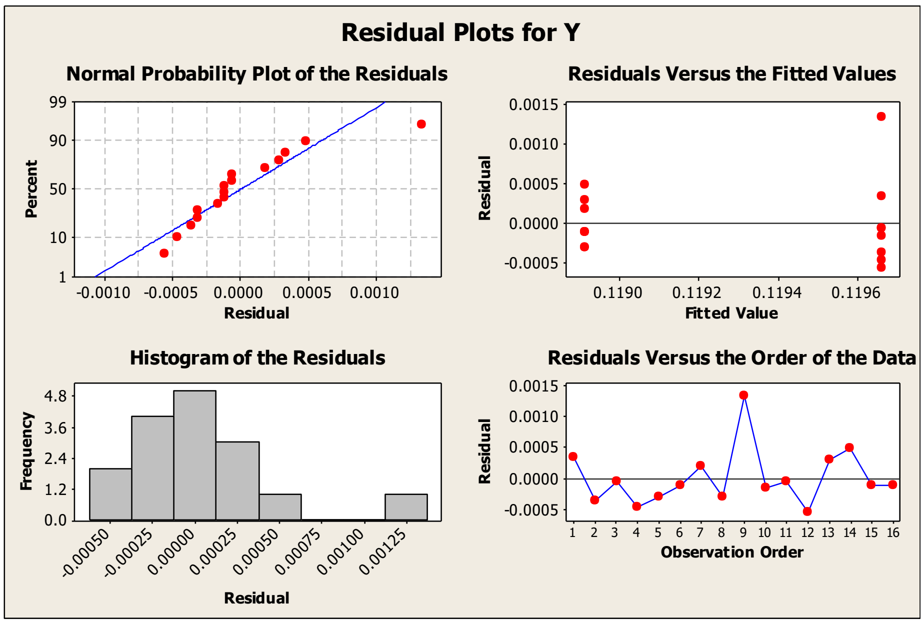
(b) Consider the following computer output. Because the *P*-value of factor B is less than $\alpha = 0.05$, we reject the null hypothesis and conclude that the main factor of factor B is significant.

Factorial Fit: Y versus B

```
Estimated Effects and Coefficients for Y (coded units)
Term         Effect       Coef    SE Coef        T       P
Constant               0.119288  0.000119   999.99   0.000
B          0.000750    0.000375  0.000119     3.14   0.007

S = 0.000477157   R-Sq = 41.38%   R-Sq(adj) = 37.19%
```

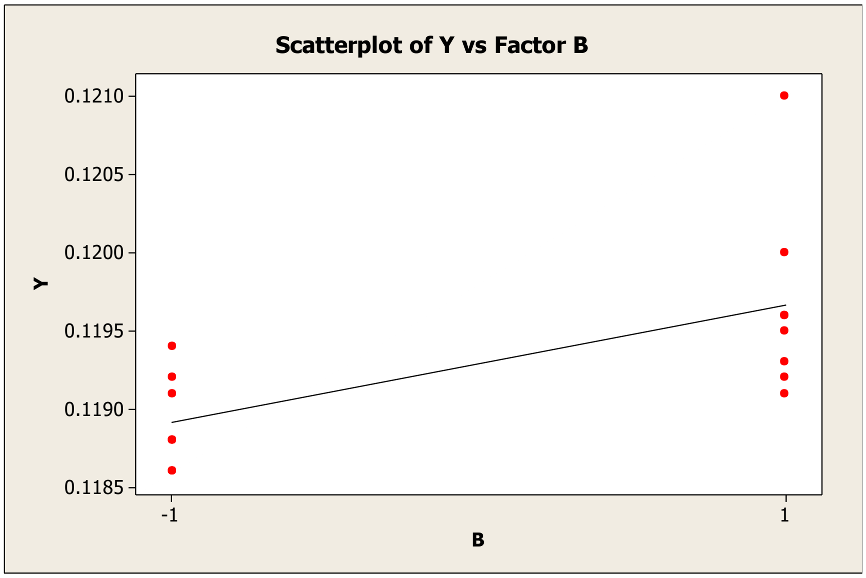
```
Analysis of Variance for Y (coded units)
Source           DF       Seq SS       Adj SS       Adj MS     F      P
Main Effects      1   0.00000225   0.00000225   0.00000225  9.88  0.007
Residual Error   14   0.00000319   0.00000319   0.00000023
  Pure Error     14   0.00000319   0.00000319   0.00000023
Total            15   0.00000544
```

(c)



The normal probability plot does not indicate any serious concerns with assumptions. The plot of residuals versus the predicted values shows a potential problem of non-constant of variance. The actual time order of the observations was not provided, so the plot versus observation order is not relevant.

(d) Only one main factor B is significant. The design is reduced to 8 replicates of an experiment with a single factor with two levels. The scatter plot of Y and factor B indicates that an increase to factor B increases the response.



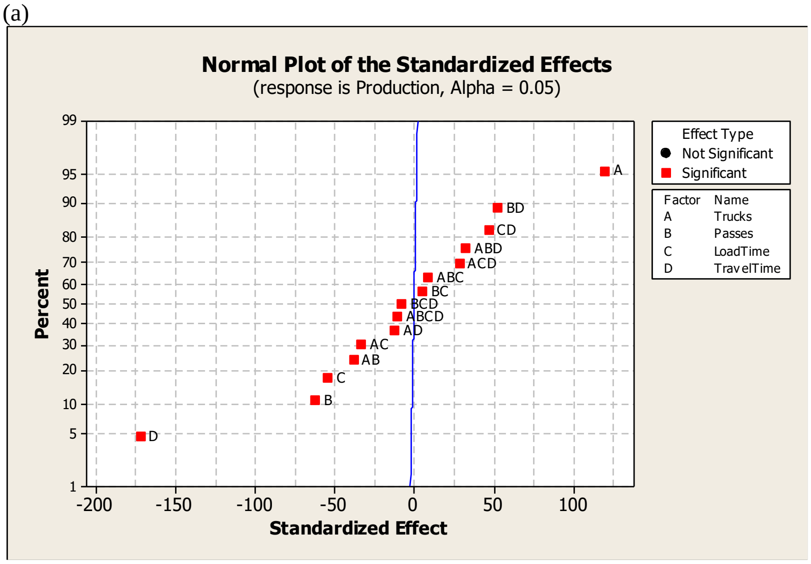
14-34 An article in *Journal of Construction Engineering and Management* ("Analysis of Earth-Moving Systems Using Discrete—Event Simulation," 1995, Vol. 121(4), pp. 388–396) considered a replicated two-level factorial experiment to study the factors most important to output in an earth-moving system. Handle the experiment as four replicates of a 2^4 factorial design with response equal to production rate (m^3 / h). The data are shown in the following table.

(a) Estimate the factor effects. Based on a normal probability plot of the effect estimates, identify a model for the data from this experiment.
(b) Conduct an ANOVA based on the model identified in part (a). What are the conclusions?
(c) Analyze the residuals and plot residuals versus the predicted production.
(d) Comment on model adequacy.

Row	Number of Trucks	Passes per Load	Load-pass Time	Travel Time	Production (m^3/h)			
					1	2	3	4
1	−	−	−	−	179.6	179.8	176.3	173.1
2	+	−	−	−	373.1	375.9	372.4	361.1
3	−	+	−	−	153.2	153.6	150.8	148.6
4	+	+	−	−	226.1	220.0	225.7	218.5
5	−	−	+	−	156.9	155.4	154.2	152.2
6	+	−	+	−	242.0	233.5	242.3	233.6
7	−	+	+	−	122.7	119.6	120.9	118.6
8	+	+	+	−	135.7	130.9	135.5	131.6
9	−	−	−	+	44.2	44.0	43.5	43.6
10	+	−	−	+	124.2	123.3	122.8	121.6
11	−	+	−	+	42.0	42.4	42.5	41.0
12	+	+	−	+	116.3	117.3	115.6	114.7
13	−	−	+	+	42.1	42.6	42.8	42.9
14	+	−	+	+	119.1	119.5	116.9	117.2
15	−	+	+	+	39.6	39.7	39.5	39.2
16	+	+	+	+	107.0	105.3	104.2	103.0

Level	−1	1
Number of trucks	2	6
Passes per load	4	7
Load pass time	12 s	22 s
Travel time	100 s	800 s

(a)




```
Estimated Effects and Coefficients for Production (coded units)

Term                                 Effect     Coef  SE Coef        T      P
Constant                                      137.39   0.3410   402.87  0.000
Trucks                                81.84    40.92   0.3410   119.98  0.000
Passes                               -42.20   -21.10   0.3410   -61.87  0.000
LoadTime                             -36.89   -18.45   0.3410   -54.09  0.000
TravelTime                          -117.31   -58.65   0.3410  -171.99  0.000
Trucks*Passes                        -25.99   -13.00   0.3410   -38.11  0.000
Trucks*LoadTime                      -22.56   -11.28   0.3410   -33.08  0.000
Trucks*TravelTime                     -8.31    -4.16   0.3410   -12.19  0.000
Passes*LoadTime                        3.44     1.72   0.3410     5.04  0.000
Passes*TravelTime                     35.89    17.94   0.3410    52.62  0.000
LoadTime*TravelTime                   31.99    16.00   0.3410    46.91  0.000
Trucks*Passes*LoadTime                 5.89     2.95   0.3410     8.64  0.000
Trucks*Passes*TravelTime              22.16    11.08   0.3410    32.48  0.000
Trucks*LoadTime*TravelTime            19.51     9.76   0.3410    28.61  0.000
Passes*LoadTime*TravelTime            -5.33    -2.66   0.3410    -7.81  0.000
Trucks*Passes*LoadTime*TravelTime     -7.16    -3.58   0.3410   -10.49  0.000


S = 2.72827       PRESS = 635.173
R-Sq = 99.92%     R-Sq(pred) = 99.86%    R-Sq(adj) = 99.90%
```

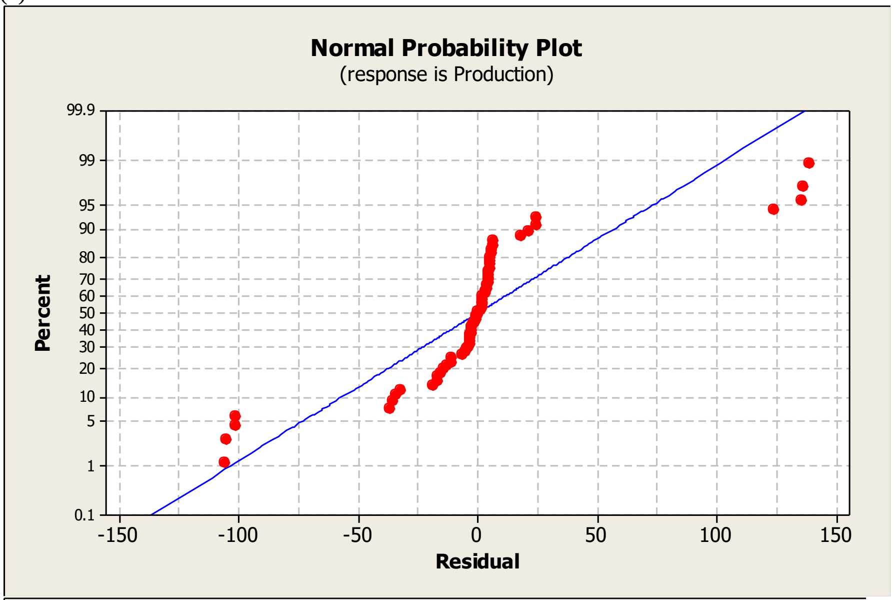
The error estimate in this experiment is small so all effects are significant. However, from the magnitude of the effects, the model is dominated by the main effects of A and D. This is consistent with the normal probability plot of the effects. Therefore, the model with only these two effects is considered.

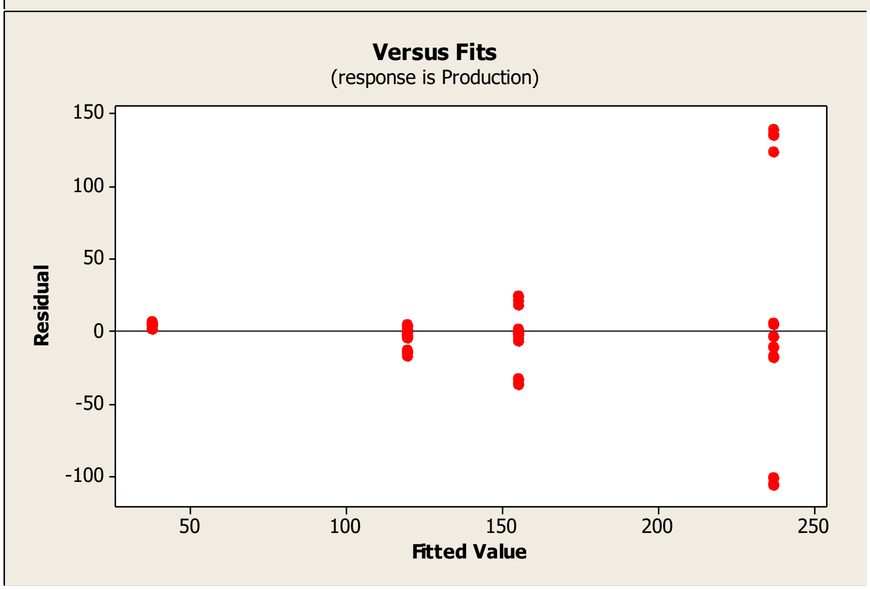
(b) An ANOVA with only the main effects of A and D follows.

```
Estimated Effects and Coefficients for Production (coded units)

Term          Effect     Coef  SE Coef       T      P
Constant               137.39    5.628   24.41  0.000
Trucks         81.84    40.92    5.628    7.27  0.000
TravelTime   -117.31   -58.65    5.628  -10.42  0.000


S = 45.0207       PRESS = 136099
R-Sq = 72.58%     R-Sq(pred) = 69.82%    R-Sq(adj) = 71.68%
```

```
Analysis of Variance for Production (coded units)

Source            DF  Seq SS  Adj SS  Adj MS       F      P
Main Effects       2  327330  327330  163665   80.75  0.000
  Trucks           1  107158  107158  107158   52.87  0.000
  TravelTime       1  220172  220172  220172  108.63  0.000
Residual Error    61  123639  123639    2027
  Lack of Fit      1    1106    1106    1106    0.54  0.465
  Pure Error      60  122533  122533    2042
Total             63  450969
```

(c)





The residuals exhibit non-constant variance. The variance increases with the fitted value. This might be expected with a measurement such as production rate.

(d) Because of the non-constant variance the model is not adequate and a transformation might be useful.

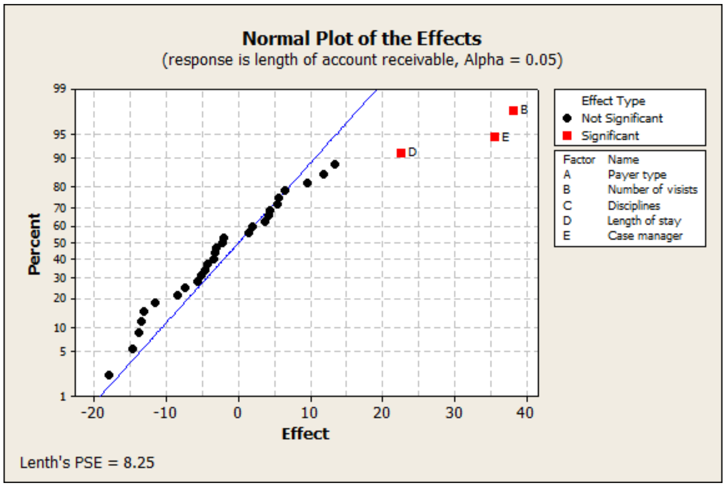
14-35 *The book Using Designed Experiments to Shrink Health Care Costs* [1997, ASQ Quality Press] presented a case study of an unreplicated 2^5 factorial design to investigate the effect of five factors on the length of accounts receivable measured in days. A summary of the investigated factors and the results of the study follows.

Row	Payer Type	Number of Visits	Disciplines	Length of Stay	Case Manager	Length of Account Receivable (days)
1	−	−	−	−	−	17
2	+	−	−	−	−	28
3	−	+	−	−	−	40
4	+	+	−	−	−	31
5	−	−	+	−	−	5
6	+	−	+	−	−	28
7	−	+	+	−	−	43
8	+	+	+	−	−	47
9	−	−	−	+	−	26
10	+	−	−	+	−	29
11	−	+	−	+	−	60
12	+	+	−	+	−	47
13	−	−	+	+	−	18
14	+	−	+	+	−	32
15	−	+	+	+	−	64
16	+	+	+	+	−	49
17	−	−	−	−	+	33
18	+	−	−	−	+	31
19	−	+	−	−	+	67
20	+	+	−	−	+	79
21	−	−	+	−	+	32
22	+	−	+	−	+	46
23	−	+	+	−	+	86
24	+	+	+	−	+	55
25	−	−	−	+	+	41
26	+	−	−	+	+	63
27	−	+	−	+	+	77
28	+	+	−	+	+	197
29	−	−	+	+	+	62
30	+	−	+	+	+	52
31	−	+	+	+	+	143
32	+	+	+	+	+	68

Level	−1	1
Payer type	Medicare	Risk Medicare
Number of visits	9	10
Disciplines	2	3c
Length of stay	30	31
Case manager	Registered Nurses	Physical Therapists

(a) Construct the normal probability plot of the effects and interpret the plot.
(b) Pool the negligible higher-order interactions to obtain an estimate of the error and construct the ANOVA accordingly.
(c) Analyze the residuals and comment on model adequacy.

(a)



From the normal plot, the main effects B, D, and E are significant effects.

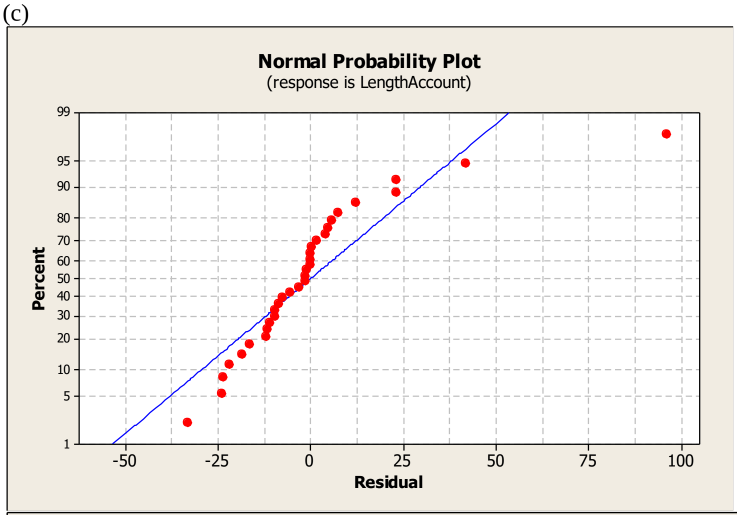
Estimated Effects and Coefficients for LengthAccount (coded units)

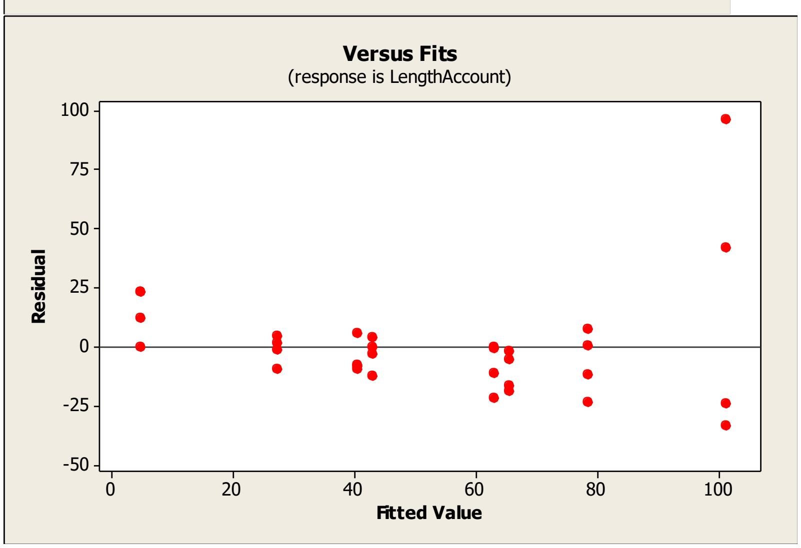
Term	Effect	Coef
Constant		53.000
PayerVisits	4.250	2.125
NumberVisits	38.125	19.062
Disciplines	-2.250	-1.125
LengthStay	22.500	11.250
CaseManager	35.500	17.750
PayerVisits*NumberVisits	-5.125	-2.563
PayerVisits*Disciplines	-13.750	-6.875
PayerVisits*LengthStay	1.500	0.750
PayerVisits*CaseManager	2.000	1.000
NumberVisits*Disciplines	-3.125	-1.562
NumberVisits*LengthStay	9.625	4.813
NumberVisits*CaseManager	13.375	6.688
Disciplines*LengthStay	-4.250	-2.125
Disciplines*CaseManager	-3.250	-1.625
LengthStay*CaseManager	11.750	5.875
PayerVisits*NumberVisits*Disciplines	-14.625	-7.312
PayerVisits*NumberVisits*LengthStay	3.625	1.812
PayerVisits*Disciplines*LengthStay	-13.500	-6.750
PayerVisits*NumberVisits*CaseManager	5.375	2.687
PayerVisits*Disciplines*CaseManager	-18.000	-9.000
PayerVisits*LengthStay*CaseManager	6.500	3.250
NumberVisits*Disciplines*LengthStay	-4.625	-2.313
NumberVisits*Disciplines*CaseManager	-8.375	-4.187
NumberVisits*LengthStay*CaseManager	5.625	2.813
Disciplines*LengthStay*CaseManager	-3.500	-1.750
PayerVisits*NumberVisits* Disciplines*LengthStay	-7.375	-3.687
PayerVisits*NumberVisits* Disciplines*CaseManager	-13.125	-6.563
PayerVisits*NumberVisits*LengthStay* CaseManager	4.375	2.188
PayerVisits*Disciplines*LengthStay* CaseManager	-11.500	-5.750
NumberVisits*Disciplines*LengthStay* CaseManager	-2.125	-1.063
PayerVisits*NumberVisits* Disciplines*LengthStay*CaseManager	-5.625	-2.812

(b)

```
Estimated Effects and Coefficients for length of account receivable (coded
units)

Term                 Effect    Coef  SE Coef      T      P
Constant                      53.00    4.275  12.40  0.000
Number of visits      38.12   19.06    4.275   4.46  0.000
Length of stay        22.50   11.25    4.275   2.63  0.014
Case manager          35.50   17.75    4.275   4.15  0.000

S = 24.1823      PRESS = 21386.3
R-Sq = 61.14%    R-Sq(pred) = 49.24%    R-Sq(adj) = 56.97%

Analysis of Variance for length of account receivable (coded units)

Source                 DF   Seq SS   Adj SS   Adj MS      F      P
Main Effects            3    25760    25760   8586.7  14.68  0.000
  Number of visists     1    11628    11628  11628.1  19.88  0.000
  Length of stay        1     4050     4050   4050.0   6.93  0.014
  Case manager          1    10082    10082  10082.0  17.24  0.000
Residual Error         28    16374    16374    584.8
  Lack of Fit           4     3530     3530    882.5   1.65  0.195
  Pure Error           24    12844    12844    535.2
Total                  31    42134
```

(c)





The normal probability plot indicates that the residuals may a heavy upper tail and the residual versus fitted plot indicates that the variance of residuals is not constant. A transformation (such as a logarithm of the length of account receivable) might be useful.

Section 14-6

14-36 Consider the data from the first replicate of Exercise 14-14.

(a) Suppose that these observations could not all be run under the same conditions. Set up a design to run these Observations in two blocks of four observations each with *ABC* confounded.
(b) Analyze the data.

(a)

BLOCK	A	B	C	y
1	-1	-1	-1	221
1	1	1	-1	552
1	1	-1	1	406
1	-1	1	1	605
2	1	-1	-1	325
2	-1	1	-1	354
2	-1	-1	1	440
2	1	1	1	392

```
Term                   Effect      Coef
Constant                         411.87
Block                             34.12
factor_A                13.75      6.87
factor_B               127.75     63.87
factor_C                97.75     48.87
factor_A*factor_B      -21.25    -10.63
factor_A*factor_C     -137.25    -68.63
factor_B*factor_C      -52.25    -26.13

Term                   Effect      Coef  StDev Coef       T      P
Constant                         411.87       19.94   20.65  0.002
Block                             34.12       19.94    1.71  0.229
factor_A                13.75      6.87       19.94    0.34  0.763
factor_B               127.75     63.87       19.94    3.20  0.085
factor_C                97.75     48.87       19.94    2.45  0.134
factor_A*factor_C     -137.25    -68.63       19.94   -3.44  0.075

Analysis of Variance for life
Source                 DF      Seq SS     Adj SS     Adj MS      F       P
Blocks                  1        9316       9316       9316   2.93   0.229
Main Effects            3       52128      52128      17376   5.46   0.159
2-Way Interactions      1       37675      37675      37675  11.84   0.075
Residual Error          2        6363       6363       3182
Total                   7      105483
```

(b) In this model with blocking, there are no significant factors.

14-37 Consider the data from the first replicate of Exercise 14-15.

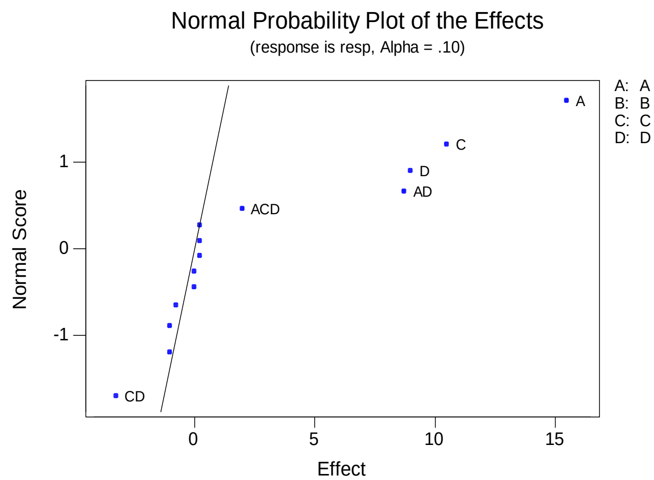
(a) Construct a design with two blocks of eight observations each with *ABCD* confounded.
(b) Analyze the data.

Design with 2 blocks

Blocks	A	B	C	D	Rep I
1	-1	-1	-1	-1	159

1	1	1	-1	-1	166
1	1	-1	1	-1	179
1	-1	1	1	-1	173
1	1	-1	-1	1	187
1	-1	1	-1	1	163
1	-1	-1	1	1	168
1	1	1	1	1	194
2	1	-1	-1	-1	168
2	-1	1	-1	-1	158
2	-1	-1	1	-1	175
2	1	1	1	-1	179
2	-1	-1	-1	1	164
2	1	1	-1	1	185
2	1	-1	1	1	197
2	-1	1	1	1	170

```
Term          Effect      Coef
Constant                174.063
Block                    -0.438
A             15.625      7.813
B             -1.125     -0.563
C             10.625      5.312
D              8.875      4.437
A*B           -0.625     -0.313
A*C            0.125      0.062
A*D            8.875      4.437
B*C            0.375      0.188
B*D            0.125      0.063
C*D           -3.125     -1.562
A*B*C         -0.125     -0.062
A*B*D         -0.875     -0.437
A*C*D          1.875      0.937
B*C*D          0.125      0.063
```



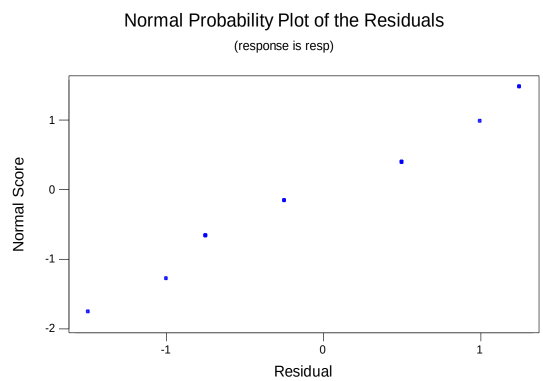
Factors A, C, and D, and interactions AD, CD and ACD appear to be significant.

```
Estimated Effects and Coefficients for resp (coded units)


Term       Effect      Coef   SE Coef        T       P
```

Constant		174.063	0.2864	607.74	0.000
Block		0.438	0.2864	1.53	0.165
A	15.625	7.812	0.2864	27.28	0.000
C	10.625	5.313	0.2864	18.55	0.000
D	8.875	4.438	0.2864	15.49	0.000
A*D	8.875	4.437	0.2864	15.49	0.000
C*D	-3.125	-1.563	0.2864	-5.46	0.001
A*C*D	1.875	0.937	0.2864	3.27	0.011

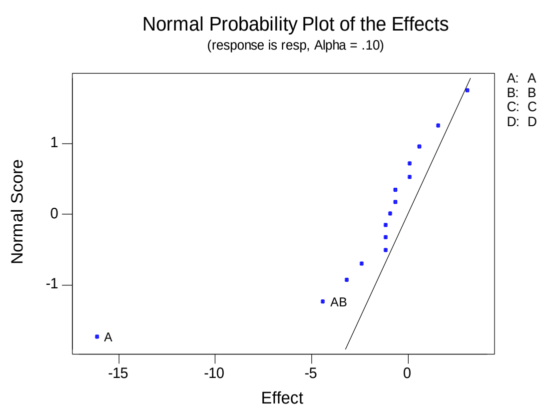
```
S = 1.14564   R-Sq = 99.51%   R-Sq(adj) = 99.07%
```

The main effects and interactions are all significant in a model that includes the factors listed above. The normal probability plot appears to support the assumption of normality.



14-38 Consider the data from Exercise 14-20.

(a) Construct the design that would have been used to run this experiment in two blocks of eight runs each.
(b) Analyze the data and draw conclusions.



Factor A and interaction AB are significant. Factor B is included in the model to make the model hierarchical.

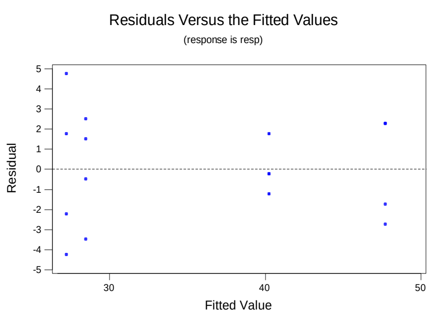
Term	Effect	Coef
Constant		35.938
BLOCK		-0.063
A	-16.125	-8.062
B	3.125	1.562
C	-1.125	-0.562
D	-1.125	-0.562
A*B	-4.375	-2.188
A*C	-0.625	-0.313

```
A*D          -3.125    -1.563
B*C           1.625     0.813
B*D           0.125     0.063
C*D          -0.625    -0.312
A*B*C         0.625     0.312
A*B*D        -2.375    -1.187
A*C*D        -1.125    -0.562
B*C*D        -0.875    -0.438

Estimated Effects and Coefficients for resp (coded units)
Term         Effect      Coef     SE Coef        T       P
Constant                35.938     0.7043    51.02   0.000
A           -16.125    -8.062      0.7043   -11.45   0.000
B             3.125     1.562      0.7043     2.22   0.047
A*B          -4.375    -2.188      0.7043    -3.11   0.009
Analysis of Variance for resp (coded units)

Source                 DF      Seq SS     Adj SS     Adj MS      F       P
Main Effects            2     1079.12    1079.12    539.562  67.98   0.000
2-Way Interactions      1       76.56      76.56     76.563   9.65   0.009
Residual Error         12       95.25      95.25      7.938
  Pure Error           12       95.25      95.25      7.938
Total                  15     1250.94
```

Effects A, B, and the AB interaction are significant at $\alpha = 0.05$. The residual analysis shows some slight differences in variability in the data.

14-39 Construct a 2^5 design in two blocks. Select the *ABCDE* interaction to be confounded with blocks.

2^5 Design in 2 Blocks with ABCDE confounded with blocks.



Run	Block	A	B	C	D	E
1	1	−	−	−	−	−
2	1	+	+	−	−	−
3	1	+	−	+	−	−
4	1	−	+	+	−	−
5	1	+	−	−	+	−
6	1	−	+	−	+	−
7	1	−	−	+	+	−
8	1	+	+	+	+	−
9	1	+	−	−	−	+
10	1	−	+	−	−	+
11	1	−	−	+	−	+
12	1	+	+	+	−	+
13	1	−	−	−	+	+
14	1	+	+	−	+	+
15	1	+	−	+	+	+
16	1	−	+	+	+	+

```
17   2    +  −  −  −  −
18   2    −  +  −  −  −
19   2    −  −  +  −  −
20   2    +  +  +  −  −
21   2    −  −  −  +  −
22   2    +  +  −  +  −
23   2    +  −  +  +  −
24   2    −  +  +  +  −
25   2    −  −  −  −  +
26   2    +  +  −  −  +
27   2    +  −  +  −  +
28   2    −  +  +  −  +
29   2    +  −  −  +  +
30   2    −  +  −  +  +
31   2    −  −  +  +  +
32   2    +  +  +  +  +
```

14-40 Consider the data from the first replicate of Exercise 14-15, assuming that four blocks are required. Confound *ABD* and *ABC* (and consequently *CD*) with blocks.

(a) Construct a design with four blocks of four observations each.
(b) Analyze the data.

(a) The design with four blocks

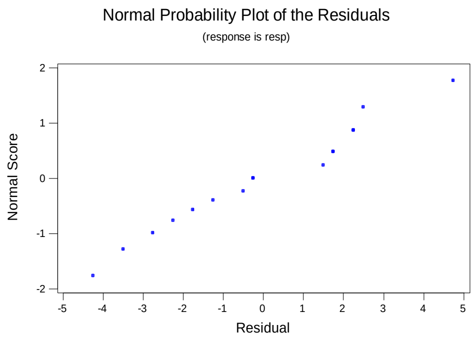
Blocks	A	B	C	D	Score
1	-1	1	1	1	170
1	1	1	-1	-1	166
1	-1	-1	-1	-1	159
1	1	-1	1	1	197
2	-1	1	1	-1	173
2	-1	-1	-1	1	164
2	1	1	-1	1	185
2	1	-1	1	-1	179
3	1	1	1	-1	179
3	1	-1	-1	1	187
3	-1	1	-1	1	163
3	-1	-1	1	-1	175
4	1	-1	-1	-1	168
4	-1	1	-1	-1	158
4	-1	-1	1	1	168
4	1	1	1	1	194

(b)

```
Estimated Effects and Coefficients for Score (coded units)

Term       Effect      Coef  SE Coef        T      P
Constant            174.063   0.5984   290.88  0.000
Block 1              -1.063   1.0364    -1.03  0.381
Block 2               1.188   1.0364     1.15  0.335
Block 3               1.937   1.0364     1.87  0.158
A          15.625     7.812   0.5984    13.06  0.001
B          -1.125    -0.563   0.5984    -0.94  0.417
C          10.625     5.313   0.5984     8.88  0.003
D           8.875     4.438   0.5984     7.42  0.005
A*B        -0.625    -0.313   0.5984    -0.52  0.638
A*C         0.125     0.063   0.5984     0.10  0.923
A*D         8.875     4.438   0.5984     7.42  0.005
B*C         0.375     0.187   0.5984     0.31  0.775
```

```
B*D         0.125    0.063   0.5984    0.10  0.923


S = 2.39357   R-Sq = 99.19%   R-Sq(adj) = 95.96%


Analysis of Variance for Score (coded units)

Source                DF   Seq SS   Adj SS   Adj MS      F      P
Blocks                 3    42.19    42.19   14.063   2.45  0.240
Main Effects           4  1748.25  1748.25  437.063  76.29  0.002
2-Way Interactions     5   317.31   317.31   63.463  11.08  0.038
Residual Error         3    17.19    17.19    5.729
Total                 15  2124.94
```

In this model with blocking there are no significant factors.



14-41 Construct a 2^5 design in four blocks. Select the appropriate effects to confound so that the highest possible interactions are confounded with blocks.

2^5 Design in 4 Blocks.

Run	Block	A	B	C	D	E
1	1	−	−	−	−	−
2	1	+	+	−	−	−
3	1	+	−	+	+	−
4	1	−	+	+	+	−
5	1	+	−	+	−	+
6	1	−	+	+	−	+
7	1	−	−	−	+	+
8	1	+	+	−	+	+
9	2	+	−	−	−	−
10	2	−	+	−	−	−
11	2	−	−	+	+	−
12	2	+	+	+	+	−
13	2	−	−	+	−	+
14	2	+	+	+	−	+
15	2	+	−	−	+	+
16	2	−	+	−	+	+
17	3	+	−	+	−	−
18	3	−	+	+	−	−
19	3	−	−	−	+	−
20	3	+	+	−	+	−
21	3	−	−	−	−	+
22	3	+	+	−	−	+
23	3	+	−	+	+	+

24	3	−	+	+	+	+
25	4	−	−	+	−	−
26	4	+	+	+	−	−
27	4	+	−	−	+	−
28	4	−	+	−	+	−
29	4	+	−	−	−	+
30	4	−	+	−	−	+
31	4	−	−	+	+	+
32	4	+	+	+	+	+

14-42 Consider the 2^6 factorial design. Set up a design to be run in four blocks of 16 runs each. Show that a design that confounds three of the four-factor interactions with blocks is the best possible blocking arrangement.

2^6 in 4 blocks.

Run	Block	A	B	C	D	E	F
1	1	−	−	−	+	+	−
2	1	+	−	+	+	−	−
3	1	−	+	−	−	+	+
4	1	−	+	+	+	−	−
5	1	+	+	−	+	+	−
6	1	−	−	+	+	+	+
7	1	+	−	+	−	+	−
8	1	+	+	+	−	−	+
9	1	−	−	+	−	−	+
10	1	+	+	−	−	−	−
11	1	−	+	+	−	+	−
12	1	+	−	−	−	+	+
13	1	−	−	−	−	−	−
14	1	−	+	−	+	−	+
15	1	+	+	+	+	+	+
16	1	+	−	−	+	−	+
17	2	−	+	+	−	+	+
18	2	+	+	+	−	−	−
19	2	−	−	−	−	−	+
20	2	−	+	+	+	−	+
21	2	+	+	−	−	−	+
22	2	+	−	−	+	−	−
23	2	−	−	−	+	+	+
24	2	−	−	+	+	+	−
25	2	−	−	+	−	−	−
26	2	+	−	+	−	+	+
27	2	−	+	−	+	−	−
28	2	+	−	+	+	−	+
29	2	+	−	−	−	+	−
30	2	+	+	+	+	+	−
31	2	−	+	−	−	+	−
32	2	+	+	−	+	+	+
33	3	+	+	+	+	−	+
34	3	−	−	−	−	+	−
35	3	−	+	+	−	−	−
36	3	−	−	+	−	+	+
37	3	−	+	−	−	−	+
38	3	−	−	+	+	−	+
39	3	+	−	−	+	+	+
40	3	+	−	+	+	+	−
41	3	+	−	+	−	−	−
42	3	+	+	−	−	+	−
43	3	−	+	+	+	+	−
44	3	−	+	−	+	+	+
45	3	+	−	−	−	−	+
46	3	+	+	−	+	−	−
47	3	−	−	−	+	−	−

```
48   3    +  +  +  −  +  +
49   4    +  −  +  +  +  +
50   4    −  −  −  −  +  +
51   4    +  −  −  −  −  −
52   4    −  −  +  +  −  −
53   4    −  +  +  +  +  +
54   4    −  +  −  −  −  −
55   4    +  −  +  −  −  +
56   4    +  +  −  +  −  +
57   4    −  +  −  +  +  −
58   4    +  +  +  −  +  −
59   4    +  +  −  −  +  +
60   4    −  −  +  −  +  −
61   4    −  +  +  −  −  +
62   4    +  −  −  +  +  −
63   4    −  −  −  +  −  +
64   4    +  +  +  +  −  −
```

14-43 An article in *Quality Engineering* ["Designed Experiment to Stabilize Blood Glucose Levels" (1999–2000, Vol. 12, pp. 83–87)] reported on an experiment to minimize variations in blood glucose levels. The factors were volume of juice intake before exercise (4 or 8 oz), amount of exercise on a Nordic Track cross-country skier (10 or 20 min), and delay between the time of juice intake (0 or 20 min) and the beginning of the exercise period. The experiment was blocked for time of day. The data follow.

(a) What effects are confounded with blocks? Comment on any concerns with the confounding in this design.
(b) Analyze the data and draw conclusions.

Run	Juice (oz)	Exercise (min)	Delay (min)	Time of Day	Average Blood Glucose
1	4	10	0	pm	71.5
2	8	10	0	am	103
3	4	20	0	am	83.5
4	8	20	0	pm	126
5	4	10	20	am	125.5
6	8	10	20	pm	129.5
7	4	20	20	pm	95
8	8	20	20	am	93

(a) Effect ABC is confounded with blocks, where A= juice, B = exercise, C = delay

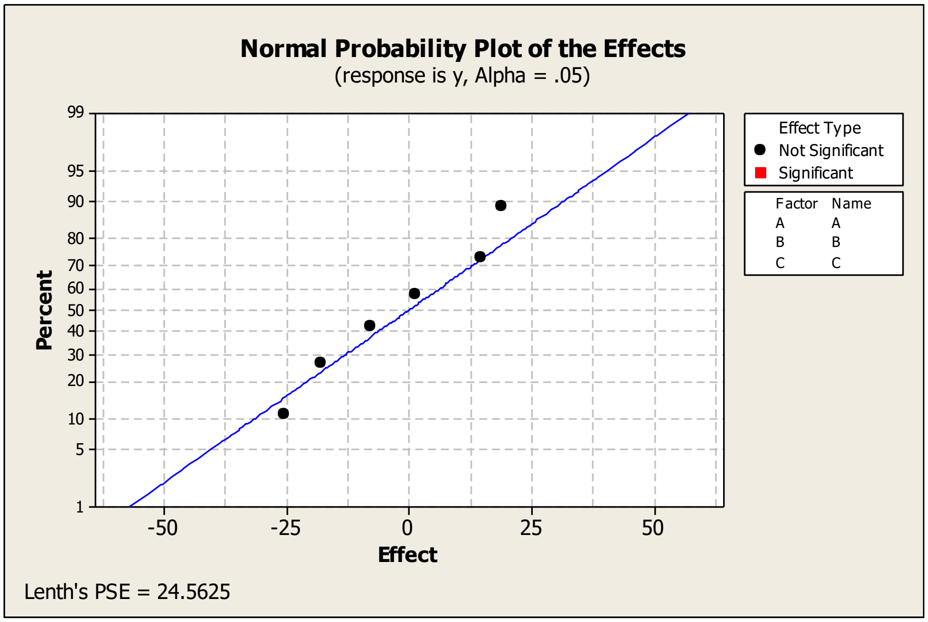
(b) In this model with blocking, there are no significant factors.

Factorial Fit: y versus Block, A, B, C

```
Estimated Effects and Coefficients for y (coded units)

Term       Effect    Coef
Constant           103.38
Block                2.13
A           19.00    9.50
B           -8.00   -4.00
C           14.75    7.38
A*B          1.25    0.62
A*C        -18.00   -9.00
B*C        -25.50  -12.75
```

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Block	36.13	1	36.13		
Model	3233.63	5	646.73	206.95	0.0527
A	722.00	1	722.00	231.04	0.0418
B	128.00	1	128.00	40.96	0.0987
C	435.13	1	435.13	139.24	0.0538
AC	648.00	1	648.00	207.36	0.0441
BC	1300.50	1	1300.50	416.16	0.0312
Residual	3.13	1	3.13		
Cor Total	3272.88	7			

Thus, the effects of *juice* as well as the interactions between *juice* and *delay* and *exercise* and *delay* were marginally significant. Additional degrees of freedom for error are needed and the normal probability plot of the effects does not indicate significant effects.



14-44 An article in *Industrial and Engineering Chemistry* ["Factorial Experiments in Pilot Plant Studies" (1951, pp. 1300–1306)] reports on an experiment to investigate the effect of temperature (*A*), gas throughput (*B*), and concentration (*C*) on the strength of product solution in a recirculation unit. Two blocks were used with *ABC* confounded, and the experiment was replicated twice. The data follow.

(a) Analyze the data from this experiment.

Replicate 1

Block 1	Block 2
(1) = 99	a = 18
ab = 52	b = 51
ac = 42	c = 108
bc = 95	abc = 35

Replicate 2

Block 4	Block 3
(1) = 46	a = 18
ab = 47	b = 62
ac = 22	c = 104
bc = 67	abc = 36

(b) Analyze the residuals and comment on model adequacy.

(c) Comment on the efficiency of this design. Note that we have replicated the experiment twice, yet we have no information on the *ABC* interaction.

(d) Suggest a better design, specifically one that would provide some information on *all* interactions.

(a) Estimated Effects and Coefficients for y

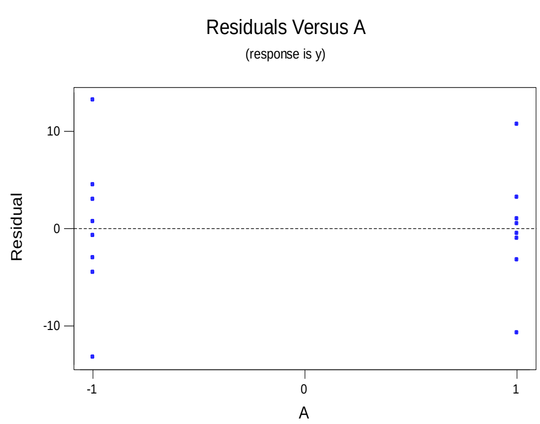
Term	Effect	Coef	StDev Coef	T	P
Constant		56.37	2.633	21.41	0.000
Block 1		15.63	4.560	3.43	0.014
2		-3.38	4.560	-0.74	0.487
3		-10.88	4.560	-2.38	0.054
A	-45.25	-22.62	2.633	-8.59	0.000
B	-1.50	-0.75	2.633	-0.28	0.785
C	14.50	7.25	2.633	2.75	0.033
A*B	19.00	9.50	2.633	3.61	0.011
A*C	-14.50	-7.25	2.633	-2.75	0.033
B*C	-9.25	-4.63	2.633	-1.76	0.130

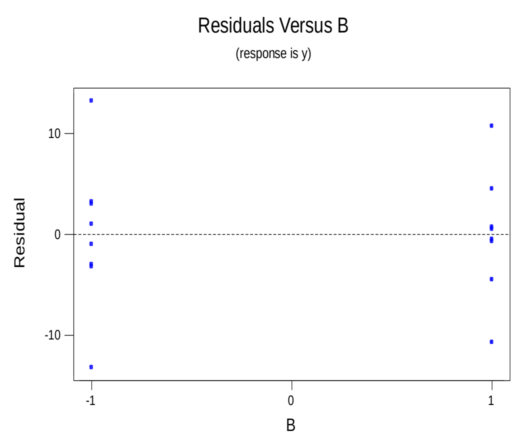
Analysis of Variance for y

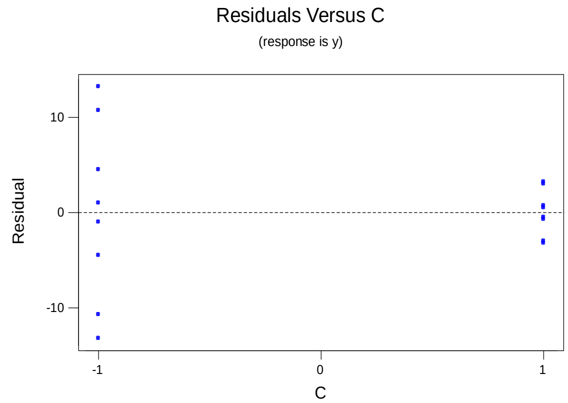
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Blocks	3	1502.8	1502.8	500.9	4.52	0.055
Main Effects	3	9040.2	9040.2	3013.4	27.17	0.001
2-Way Interactions	3	2627.2	2627.2	875.7	7.90	0.017
Residual Error	6	665.5	665.5	110.9		
Total	15	13835.7				

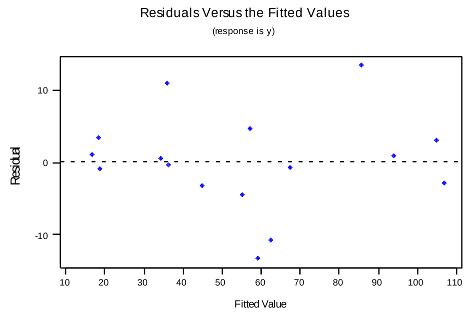
Factors A, C, AB, and AC are significant.

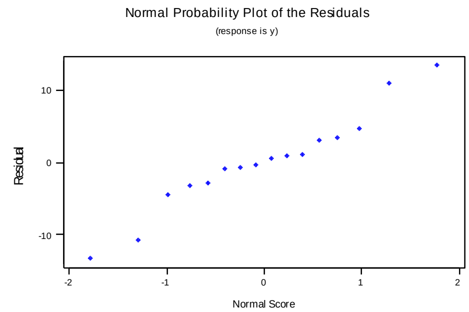
(b) There is more variability on the response associated with the low setting of factor C.











(c) Some of the information from the experiment is lost because the design is run in 4 blocks. This causes us to lose information on the ABC interaction even though we have replicated the experiment twice. If it is possible to run the experiment in only 2 blocks, there would be information on all interactions.

(d) To have data on all interactions, we could run the experiment so that each replicate is a block. In that case, there would be only two blocks.

14-45 Consider the following computer output from a single replicate of a 2^4 experiment in two blocks with ABCD confounded.

(a) Comment on the value of blocking in this experiment.
(b) What effects were used to generate the residual error in the ANOVA?
(c) Calculate the entries marked with "?" in the output.

Factorial Fit: y Versus Block, A, B, C, D
Estimated Effects and Coefficients

Term	Effect	Coef	SE Coef	*t*	P
Constant		579.33	9.928	58.35	0.000
Block		105.68	9.928	10.64	0.000
A	–15.41	–7.70	9.928	–0.78	0.481
B	2.95	1.47	9.928	0.15	0.889
C	15.92	7.96	9.928	0.80	0.468
D	–37.87	–18.94	9.928	–1.91	0.129
A*B	–8.16	–4.08	9.928	–0.41	0.702
A*C	5.91	2.95	9.928	0.30	0.781
A*D	30.28	?	9.928	?	0.202
B*C	20.43	10.21	9.928	1.03	0.362
B*D	–17.11	–8.55	9.928	–0.86	0.437
C*D	4.41	2.21	9.928	0.22	0.835

S = 39.7131 R-Sq = 96.84% R-Sq (adj) = 88.16%

(a) Because the sum of squares associated with blocks is large relative to sum of squares for residual error, we conclude that blocking is important to reduce nuisance variation in this experiment.

(b) The effects for all three-factor interaction terms (ABC, ABD, ACD, and BCD) are used to generate the residual error in ANOVA because these effects do not appear in the ANOVA table. The four-factor interaction effect is confounded with blocks.

(c) Coef of AD $= \frac{Effect}{2} = \frac{30.28}{2} = 15.14$

t test of AD $= \frac{Coef}{Se\,Coef} = \frac{15.14}{9.928} = 1.525$

The degrees of freedom for blocks are 15 – 4 – 6 – 4 = 1.
Also, we know there are two blocks so the degrees of freedom = 2 – 1 = 1.

Adj MS of 2-way Interactions $= \frac{Adj\ SS}{DF} = \frac{6992}{6} = 1165.33$

14-46 An article in *Advanced Semiconductor Manufacturing Conference (ASMC)* (May 2004, pp. 325–29) stated that dispatching rules and rework strategies are two major operational elements that impact productivity in a semiconductor fabrication plant (fab). A four-factor experiment was conducted to determine the effect of dispatching rule time (5 or 10 min), rework delay (0 or 15 min), fab temperature (60 or 80°F), and rework levels (level 0 or level 1) on key fab performance measures. The performance measure that was analyzed was the average cycle time. The experiment was blocked for the fab temperature. Data modified from the original study are in the following table.

Run	Dispatching Rule Time (min)	Rework Delay (min)	Rework Level	Fab Temperature (°F)	Average Cycle TimeRun (min)
1	5	0	0	60	218
2	10	0	0	80	256.5
3	5	0	1	80	231
4	10	0	1	60	302.5
5	5	15	0	80	298.5
6	10	15	0	60	314
7	5	15	1	60	249
8	10	15	1	80	241

(a) What effects are confounded with blocks? Do you find any concerns with confounding in this design? If so, comment on it.
(b) Analyze the data and draw conclusions.

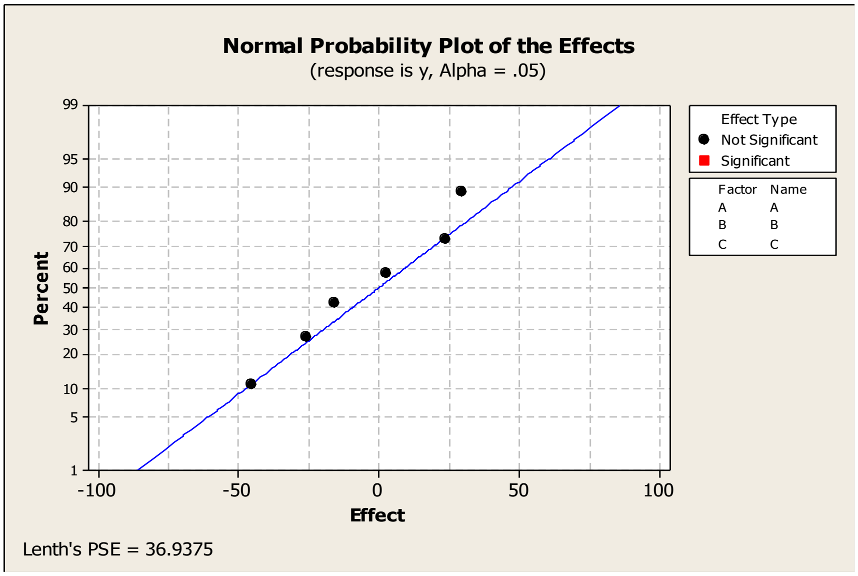
(a) The effect of Fab Temperature (D) is aliased with the three factor interaction Dispatching time (A)**Rework Delay time* (B)**Rework Level* (C) and this alias set is confounded with blocks.

That is, Block = *Fab Temperature* = *Dispatching time*Rework Delay time*Rework Level*

A concern is that the alias set confounded with blocks contains the main effect of Fab Temperature.

(b) Computer software will often not analyze an experiment with a main effect confounded with blocks. Therefore, the experiment is handled as a three-factor experiment in factors A, B, C confounded in two blocks. Information on factor D is lost because it is confounded with blocks.

```
Factorial Fit: y versus Block, A, B, C

Estimated Effects and Coefficients for y (coded units)

Term       Effect    Coef
Constant           263.81
Block                7.06
A           29.37   14.69
B           23.63   11.81
C          -15.88   -7.94
A*B        -25.63  -12.81
A*C          2.38    1.19
B*C        -45.38  -22.69
```



From the normal probability plot of effects, there does not appear to be any significant effects. However, the effect estimates in the table show that the A*C effect = 0.13 is much smaller than the others. If only this effect represents the magnitude of noise, the following computer output shows that all the other effects are significant. Some additional data is needed here to estimate noise and to choose among the results that all but the A*C effect are significant or no effects are significant.

Factorial Fit: y versus Block, A, B, C

Estimated Effects and Coefficients for y (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		263.81	1.188	222.16	0.003
Block		7.06	1.188	5.95	0.106
A	29.37	14.69	1.188	12.37	0.051
B	23.63	11.81	1.188	9.95	0.064
C	-15.88	-7.94	1.188	-6.68	0.095
A*B	-25.63	-12.81	1.188	-10.79	0.059
B*C	-45.37	-22.69	1.188	-19.11	0.033

S = 3.35876   R-Sq = 99.88%   R-Sq(adj) = 99.14%

Analysis of Variance for y (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Blocks	1	399.03	399.03	399.03	35.37	0.106
Main Effects	3	3346.09	3346.09	1115.36	98.87	0.074
2-Way Interactions	2	5431.06	5431.06	2715.53	240.71	0.046
Residual Error	1	11.28	11.28	11.28		
Total	7	9187.47				

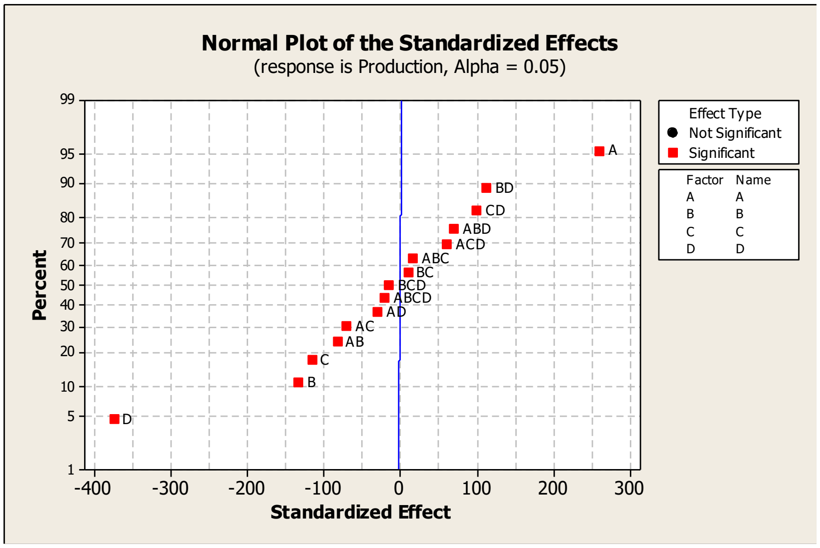
14-47 Consider the earth-moving experiment in Exercise 14-34. The experiment actually used two different operators with the production in columns 1 and 3 from operator 1 and 2, respectively. Analyze the results from only columns 1 and 3 handled as blocks.

(a) Assuming that the operator is a **nuisance factor**, estimate the factor effects.
(b) Based on a normal probability plot of the effect estimates, identify a model for the data from this experiment.
(c) Conduct an ANOVA based on the model identified in part (a). What are the conclusions?
(d) Analyze the residuals and plot residuals versus the predicted production.
(e) Compare the results from this analysis to the previous analysis that did not use blocking.

(a) Assume operator is a nuisance factor and estimate the factor effects.

```
Estimated Effects and Coefficients for Production (coded units)

Term                                 Effect     Coef  SE Coef        T      P
Constant                                      138.43   0.1591   870.00  0.000
Block                                           0.56   0.1591     3.52  0.003
NumberTrucks                          83.01    41.50   0.1591   260.84  0.000
Passes                               -42.19   -21.10   0.1591  -132.59  0.000
LoadTime                             -36.68   -18.34   0.1591  -115.27  0.000
TravelTime                          -119.07   -59.53   0.1591  -374.16  0.000
NumberTrucks*Passes                  -26.14   -13.07   0.1591   -82.15  0.000
NumberTrucks*LoadTime                -22.51   -11.25   0.1591   -70.72  0.000
NumberTrucks*TravelTime               -9.27    -4.63   0.1591   -29.13  0.000
Passes*LoadTime                        3.29     1.65   0.1591    10.35  0.000
Passes*TravelTime                     36.08    18.04   0.1591   113.38  0.000
LoadTime*TravelTime                   31.69    15.85   0.1591    99.59  0.000
NumberTrucks*Passes*LoadTime           5.57     2.78   0.1591    17.50  0.000
NumberTrucks*Passes*TravelTime        22.28    11.14   0.1591    70.02  0.000
NumberTrucks*LoadTime*TravelTime      19.57     9.78   0.1591    61.49  0.000
Passes*LoadTime*TravelTime            -4.83    -2.42   0.1591   -15.18  0.000
NumberTrucks*Passes*LoadTime*         -6.46    -3.23   0.1591   -20.29  0.000
  TravelTime

S = 0.900081     PRESS = 55.3060
R-Sq =  99.99%   R-Sq(pred) = 99.98%   R-Sq(adj) = 99.99%

Analysis of Variance for Production (coded units)

Source                                  DF  Seq SS  Adj SS  Adj MS
Blocks                                   1      10      10      10
Main Effects                             4  193546  193546   48386
  NumberTrucks                           1   55120   55120   55120
  Passes                                 1   14243   14243   14243
  LoadTime                               1   10764   10764   10764
  TravelTime                             1  113419  113419  113419
2-Way Interactions                       6   28745   28745    4791
  NumberTrucks*Passes                    1    5468    5468    5468
  NumberTrucks*LoadTime                  1    4052    4052    4052
  NumberTrucks*TravelTime                1     687     687     687
  Passes*LoadTime                        1      87      87      87
  Passes*TravelTime                      1   10415   10415   10415
  LoadTime*TravelTime                    1    8036    8036    8036
3-Way Interactions                       4    7470    7470    1867
  NumberTrucks*Passes*LoadTime           1     248     248     248
  NumberTrucks*Passes*TravelTime         1    3972    3972    3972
  NumberTrucks*LoadTime*TravelTime       1    3063    3063    3063
  Passes*LoadTime*TravelTime             1     187     187     187
4-Way Interactions                       1     333     333     333
  NumberTrucks*Passes*LoadTime*TravelTime  1   333     333     333
Residual Error                          15      12      12       1
Total                                   31  230117
```

(b)



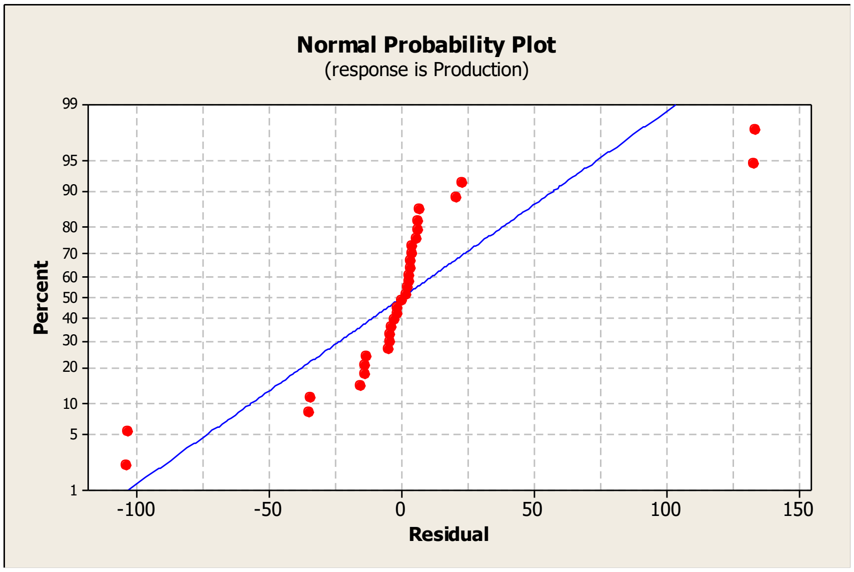
The error term is small so all the effects are significant. However, the main effects of A and D are much greater than the others.

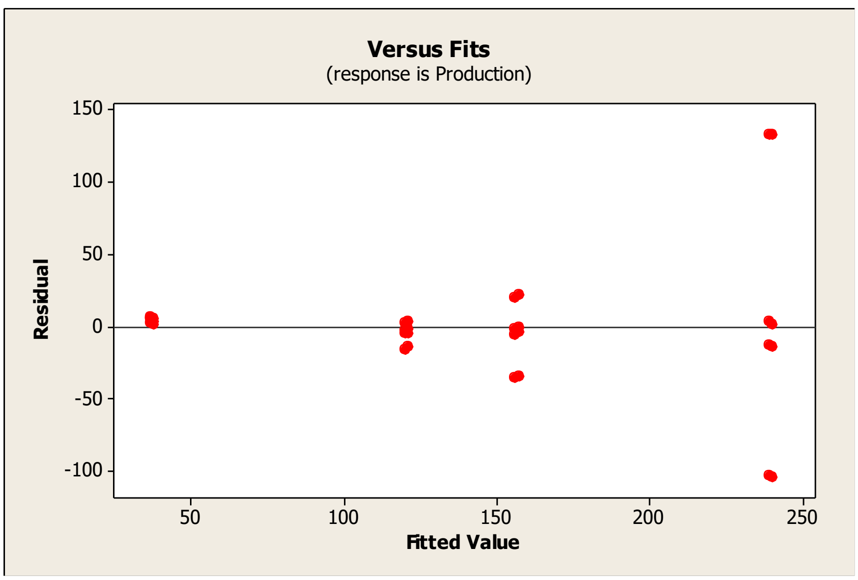
(c)

```
Estimated Effects and Coefficients for Production (coded units)

Term            Effect    Coef  SE Coef      T      P
Constant                138.43    8.289  16.70  0.000
Block                     0.56    8.289   0.07  0.947
NumberTrucks     83.01   41.50    8.289   5.01  0.000
TravelTime     -119.07  -59.53    8.289  -7.18  0.000


S = 46.8917      PRESS = 80414.4
R-Sq = 73.25%    R-Sq(pred) = 65.05%   R-Sq(adj) = 70.38%


Analysis of Variance for Production (coded units)

Source          DF  Seq SS  Adj SS  Adj MS      F      P
Blocks           1      10      10      10   0.00  0.947
Main Effects     2  168539  168539   84270  38.32  0.000
  NumberTrucks   1   55120   55120   55120  25.07  0.000
  TravelTime     1  113419  113419  113419  51.58  0.000
Residual Error  28   61567   61567    2199
  Lack of Fit    4     697     697     174   0.07  0.991
  Pure Error    24   60871   60871    2536
Total           31  230117
```

The main effects of *NumberTrucks* and *TravelTime* are significant.

(d)





The residuals exhibit non-constant variance. The variance increases with the fitted value. This might be expected with a measurement such as production rate. Because of the non-constant variance the model is not adequate and a transformation might be useful.

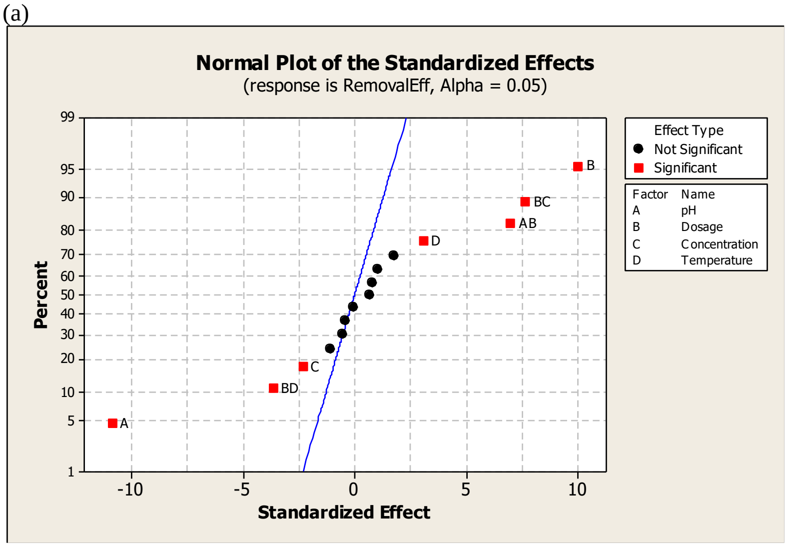
(e) The results from this analysis are similar to the ones from the unblocked analysis in a previous exercise.

14-48 An article in *Journal of Hazardous Materials* [“Biosorption of Reactive Dye Using Acid-Treated Rice Husk: Factorial Design Analysis” (2007, Vol. 142(1), pp. 397–403)] described an experiment using biosorption to remove red color from water. A 2^4 full factorial design was used to study the effect of factors pH, temperature, adsorbent dosage, and initial concentration of the dye. Handle columns 1 and 2 of the output as blocks.

(a) Estimate the factor effects. Based on a normal probability plot of the effect estimates, identify a model for the data from this experiment.
(b) Conduct an ANOVA based on the model identified in part (a). What are the conclusions?
(c) Analyze the residuals and plot residuals versus the predicted removal efficiency.
(d) Construct a regression model to predict removal efficiency in terms of the actual factor levels.

Run	pH	Dosage (g/L)	Concentration (mg/L)	Temperature (°C)	Removal Efficiency (%)	
					1	2
1	2	5	50	20	89.36	95.78
2	7	5	50	20	53.67	52.02
3	2	50	50	20	86.97	93.76
4	7	50	50	20	72.39	80.55
5	2	5	250	20	68.46	64.99
6	7	5	250	20	32.44	28.44
7	2	50	250	20	93.19	93.69
8	7	50	250	20	88.17	91.41
9	2	5	50	40	97.25	95.41
10	7	5	50	40	76.42	56.51
11	2	50	50	40	76.24	90.83
12	7	50	50	40	79.54	73.21
13	2	5	250	40	84.31	82.84
14	7	5	250	40	53.32	44.96
15	2	50	250	40	94.77	96.53
16	7	50	250	40	89.32	90.75

(a)




```
Estimated Effects and Coefficients for RemovalEff (coded units)

Term                          Effect    Coef  SE Coef       T      P
Constant                               77.11   0.9812   78.59  0.000
Block                                   0.13   0.9812    0.13  0.897
pH                            -21.33  -10.66   0.9812  -10.87  0.000
Dosage                         19.70    9.85   0.9812   10.04  0.000
Concentration                  -4.52   -2.26   0.9812   -2.30  0.036
Temperature                     6.06    3.03   0.9812    3.09  0.008
pH*Dosage                      13.75    6.87   0.9812    7.01  0.000
pH*Concentration                1.33    0.67   0.9812    0.68  0.507
pH*Temperature                  2.06    1.03   0.9812    1.05  0.310
Dosage*Concentration           15.06    7.53   0.9812    7.68  0.000
Dosage*Temperature             -7.18   -3.59   0.9812   -3.66  0.002
Concentration*Temperature       3.44    1.72   0.9812    1.75  0.100
pH*Dosage*Concentration         1.61    0.81   0.9812    0.82  0.423
pH*Dosage*Temperature          -0.87   -0.43   0.9812   -0.44  0.665
```

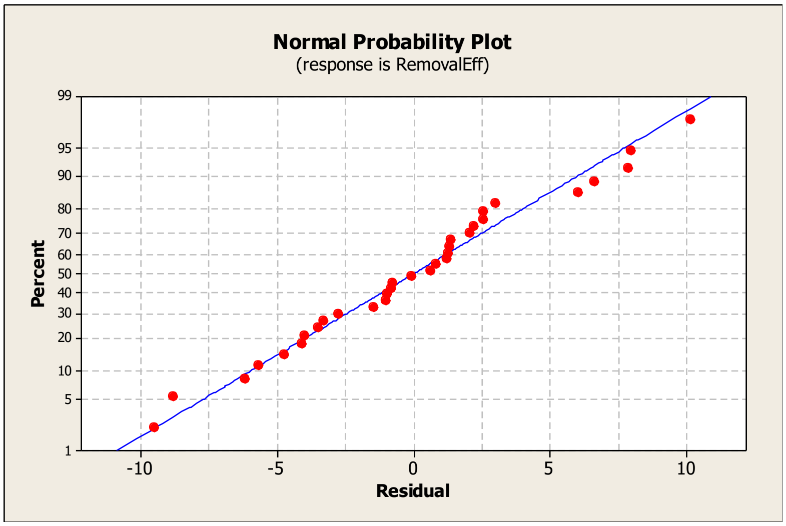
```
pH*Concentration*Temperature          -2.09   -1.04   0.9812   -1.06  0.304
Dosage*Concentration*Temperature      -1.10   -0.55   0.9812   -0.56  0.584
pH*Dosage*Concentration*Temperature   -0.09   -0.04   0.9812   -0.04  0.966


S = 5.55058     PRESS = 2103.23
R-Sq = 96.02%   R-Sq(pred) = 81.89%   R-Sq(adj) = 91.78%


Analysis of Variance for RemovalEff (coded units)

Source                                DF   Seq SS   Adj SS   Adj MS       F
Blocks                                 1      0.5     0.54     0.54    0.02
Main Effects                           4   7199.9  7199.85  1799.96   58.42
  pH                                   1   3639.3  3639.32  3639.32  118.13
  Dosage                               1   3103.5  3103.54  3103.54  100.73
  Concentration                        1    163.4   163.44   163.44    5.31
  Temperature                          1    293.5   293.55   293.55    9.53
2-Way Interactions                     6   3882.1  3882.13   647.02   21.00
  pH*Dosage                            1   1512.2  1512.23  1512.23   49.08
  pH*Concentration                     1     14.2    14.20    14.20    0.46
  pH*Temperature                       1     33.9    33.95    33.95    1.10
  Dosage*Concentration                 1   1815.0  1815.03  1815.03   58.91
  Dosage*Temperature                   1    411.8   411.85   411.85   13.37
  Concentration*Temperature            1     94.9    94.88    94.88    3.08
3-Way Interactions                     4     71.4    71.45    17.86    0.58
  pH*Dosage*Concentration              1     20.9    20.87    20.87    0.68
  pH*Dosage*Temperature                1      6.0     6.02     6.02    0.20
  pH*Concentration*Temperature         1     34.9    34.90    34.90    1.13
  Dosage*Concentration*Temperature     1      9.7     9.66     9.66    0.31
4-Way Interactions                     1      0.1     0.06     0.06    0.00
  pH*Dosage*Concentration*Temperature  1      0.1     0.06     0.06    0.00
Residual Error                        15    462.1   462.13    30.81
Total                                 31  11616.2

Source                                    P
Blocks                                0.897
Main Effects                          0.000
  pH                                  0.000
  Dosage                              0.000
  Concentration                       0.036
  Temperature                         0.008
2-Way Interactions                    0.000
  pH*Dosage                           0.000
  pH*Concentration                    0.507
  pH*Temperature                      0.310
  Dosage*Concentration                0.000
  Dosage*Temperature                  0.002
  Concentration*Temperature           0.100
3-Way Interactions                    0.682
  pH*Dosage*Concentration             0.423
  pH*Dosage*Temperature               0.665
  pH*Concentration*Temperature        0.304
  Dosage*Concentration*Temperature    0.584
4-Way Interactions                    0.966
  pH*Dosage*Concentration*Temperature 0.966
Residual Error
Total
```

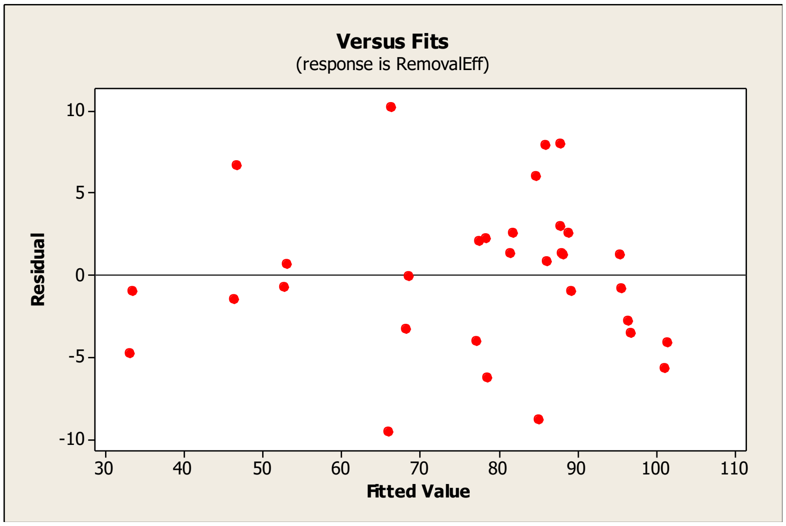
The main effects of *pH*, *Dosage*, *Concentration*, and *Temperature* and the interactions *pH*Dosage*, *Dosage*Concentration*, and *Dosage*Temperature* are significant.

(b) Using only the significant effects

```
Estimated Effects and Coefficients for RemovalEff (coded units)

Term                    Effect    Coef  SE Coef      T      P
Constant                         77.11   0.9588  80.42  0.000
Block                             0.13   0.9588   0.13  0.894
pH                      -21.33  -10.66   0.9588 -11.12  0.000
Dosage                   19.70    9.85   0.9588  10.27  0.000
Concentration            -4.52   -2.26   0.9588  -2.36  0.027
Temperature               6.06    3.03   0.9588   3.16  0.004
pH*Dosage                13.75    6.87   0.9588   7.17  0.000
Dosage*Concentration     15.06    7.53   0.9588   7.85  0.000
Dosage*Temperature       -7.18   -3.59   0.9588  -3.74  0.001


S = 5.42406      PRESS = 1309.85
R-Sq = 94.17%    R-Sq(pred) = 88.72%    R-Sq(adj) = 92.15%

Analysis of Variance for RemovalEff (coded units)

Source                  DF   Seq SS   Adj SS   Adj MS       F      P
Blocks                   1      0.5     0.54     0.54    0.02  0.894
Main Effects             4   7199.9  7199.85  1799.96   61.18  0.000
  pH                     1   3639.3  3639.32  3639.32  123.70  0.000
  Dosage                 1   3103.5  3103.54  3103.54  105.49  0.000
  Concentration          1    163.4   163.44   163.44    5.56  0.027
  Temperature            1    293.5   293.55   293.55    9.98  0.004
2-Way Interactions       3   3739.1  3739.10  1246.37   42.36  0.000
  pH*Dosage              1   1512.2  1512.23  1512.23   51.40  0.000
  Dosage*Concentration   1   1815.0  1815.03  1815.03   61.69  0.000
  Dosage*Temperature     1    411.8   411.85   411.85   14.00  0.001
Residual Error          23    676.7   676.67    29.42
Total                   31  11616.2
```

(c) The residual plots for the reduced model follow. The residual plots do not indicate departures from the assumptions.





(d) The block effect is small and can be ignored.

```
Estimated Coefficients for RemovalEff using data in uncoded units

Term                          Coef
Constant                   94.3492
Block                     0.129375
pH                        -7.62656
Dosage                   -0.136006
Concentration            -0.114649
Temperature               0.741347
pH*Dosage                 0.122211
Dosage*Concentration    0.00334722
Dosage*Temperature      -0.0159444
```

Section 14-7

14-49 Consider the problem in Exercise 14-19. Suppose that only half of the 32 runs could be made.

(a) Choose the half that you think should be run.
(b) Write out the alias relationships for your design.
(c) Estimate the factor effects.
(d) Plot the effect estimates on normal probability paper and interpret the results.
(e) Set up an analysis of variance for the factors identified as potentially interesting from the normal probability plot in part (d).
(f) Analyze the residuals from the model.
(g) Provide a practical interpretation of the results.

(a) Design 2^{5-1}

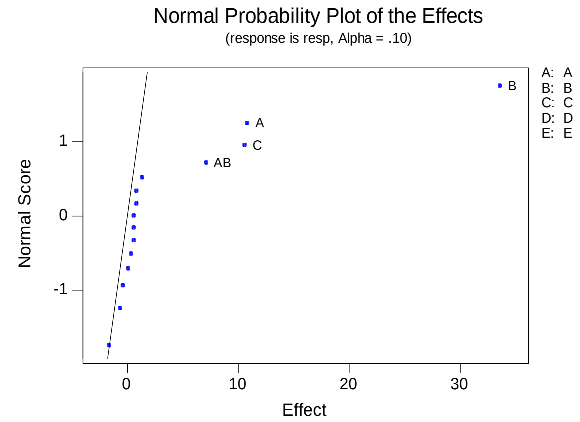
Run No.	A	B	C	D	E	resp
1	-1	-1	-1	-1	1	8
2	1	-1	-1	-1	-1	9
3	-1	1	-1	-1	-1	34
4	1	1	-1	-1	1	52
5	-1	-1	1	-1	-1	16
6	1	-1	1	-1	1	22
7	-1	1	1	-1	1	45

8	1	1	1	-1	-1	60
9	-1	-1	-1	1	-1	8
10	1	-1	-1	1	1	10
11	-1	1	-1	1	1	30
12	1	1	-1	1	-1	50
13	-1	-1	1	1	1	15
14	1	-1	1	1	-1	21
15	-1	1	1	1	-1	44
16	1	1	1	1	1	63

(b)
```
Design Generators:  E = ABCD

Alias Structure

I + ABCDE

A + BCDE
B + ACDE
C + ABDE
D + ABCE
E + ABCD
AB + CDE
AC + BDE
AD + BCE
AE + BCD
BC + ADE
BD + ACE
BE + ACD
CD + ABE
CE + ABD
DE + ABC
```

(c)
```
Term                     Effect       Coef
Constant                           30.4375
factor_A                10.8750     5.4375
factor_B                33.6250    16.8125
factor_C                10.6250     5.3125
factor_D                -0.6250    -0.3125
factor_E                 0.3750     0.1875
factor_A*factor_B        7.1250     3.5625
factor_A*factor_C        0.6250     0.3125
factor_A*factor_D        0.8750     0.4375
factor_A*factor_E        1.3750     0.6875
factor_B*factor_C        0.8750     0.4375
factor_B*factor_D       -0.3750    -0.1875
factor_B*factor_E        0.1250     0.0625
factor_C*factor_D        0.6250     0.3125
factor_C*factor_E        0.6250     0.3125
factor_D*factor_E       -1.6250    -0.8125
```

(d) Factors A, B, and C and interaction AB are significant



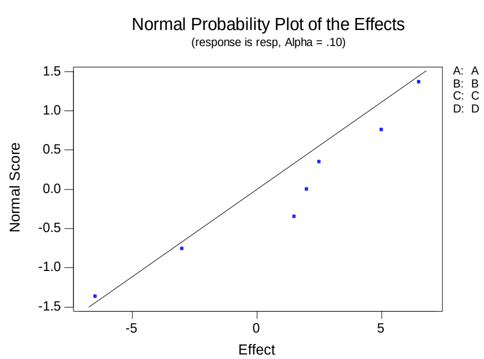
(e)

Term	Effect	Coef	StDev Coef	T	P
Constant		30.438	0.4243	71.73	0.000
factor_A	10.875	5.438	0.4243	12.81	0.000
factor_B	33.625	16.812	0.4243	39.62	0.000
factor_C	10.625	5.313	0.4243	12.52	0.000
factor_A*factor_B	7.125	3.562	0.4243	8.40	0.000

14-50 Suppose that in Exercise 14-22 it was possible to run only a $\frac{1}{2}$ fraction of the 2^4 design. Construct the design and use only the data from the eight runs you have generated to perform the analysis.

```
Estimated Effects and Coefficients for resp (coded units)

Estimated Effects and Coefficients for yield (coded units)
```

Term	Effect	Coef
Constant		17.875
A	3.750	1.875
B	0.250	0.125
C	2.750	1.375
D	4.250	2.125
A*B	-0.750	-0.375
A*C	-4.250	-2.125
A*D	4.250	2.125



None of the factors appear to be significant in the 2^{4-1} design.

14-51 An article by L. B. Hare ["In the Soup: A Case Study to Identify Contributors to Filling Variability," *Journal of Quality Technology* 1988 (Vol. 20, pp. 36–43)] described a factorial experiment used to study fi lling variability of dry soup mix packages. The factors are *A* = number of mixing ports through which the vegetable oil was added (1, 2), *B* = temperature surrounding the mixer (cooled, ambient), *C* = mixing time (60, 80 sec), *D* = batch weight (1500, 2000 lb), and *E* = number of days of delay between mixing and packaging (1, 7). Between 125 and 150 packages of soup were sampled over an eight-hour period for each run in the design, and the standard deviation of package weight was used as the response variable. The design and resulting data follow.

(a) What is the generator for this design?
(b) What is the resolution of this design?
(c) Estimate the factor effects. Which effects are large?
(d) Does a residual analysis indicate any problems with the underlying assumptions?
(e) Draw conclusions about this filling process.

Std Order	*A* Mixer Ports	*B* Temp	*C* Time	*D* Batch Weight	*E* Delay	*y* Std Dev
1	–	–	–	–	–	1.13
2	+	–	–	–	+	1.25
3	–	+	–	–	+	0.97
4	+	+	–	–	–	1.70
5	–	–	+	–	+	1.47
6	+	–	+	–	–	1.28
7	–	+	+	–	–	1.18
8	+	+	+	–	+	0.98
9	–	–	–	+	+	0.78
10	+	–	–	+	–	1.36
11	–	+	–	+	–	1.85
12	+	+	–	+	+	0.62
13	–	–	+	+	–	1.09
14	+	–	+	+	+	1.10
15	–	+	+	+	+	0.76
16	+	+	+	+	–	2.10

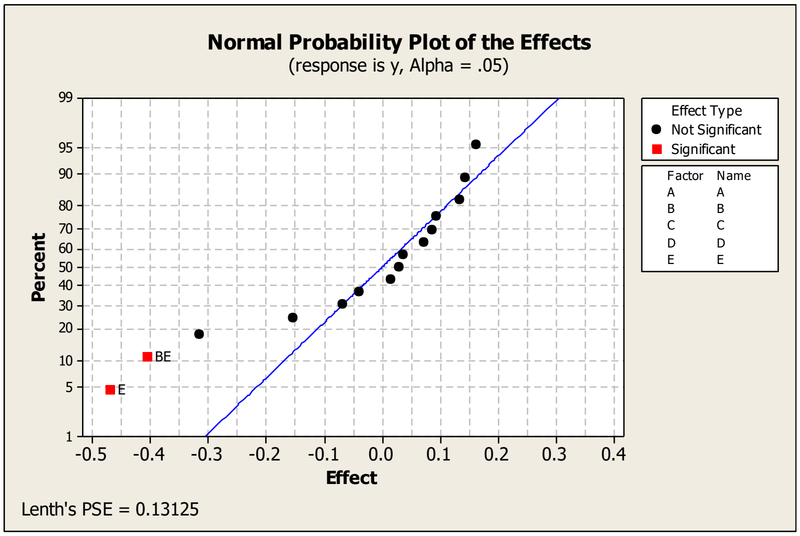
(a) The generator is E= -ABCD

(b) The resolution is resolution V

(c) `Estimated Effects and Coefficients for y (coded units)`

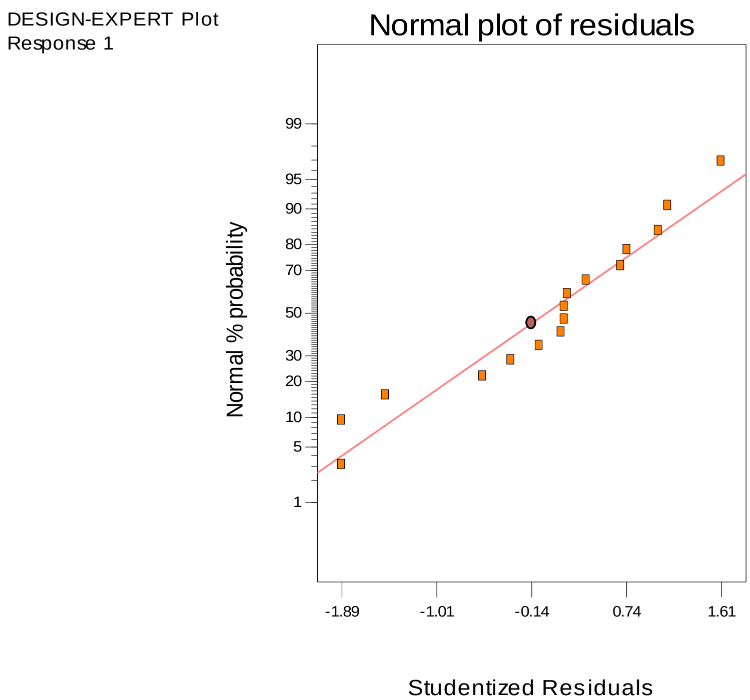
```
Term        Effect     Coef
Constant             1.2263
A           0.1450   0.0725
B           0.0875   0.0438
C           0.0375   0.0187
D          -0.0375  -0.0187
E          -0.4700  -0.2350
A*B         0.0150   0.0075
A*C         0.0950   0.0475
A*D         0.0300   0.0150
A*E        -0.1525  -0.0762
B*C        -0.0675  -0.0338
B*D         0.1625   0.0813
B*E        -0.4050  -0.2025
C*D         0.0725   0.0363
```

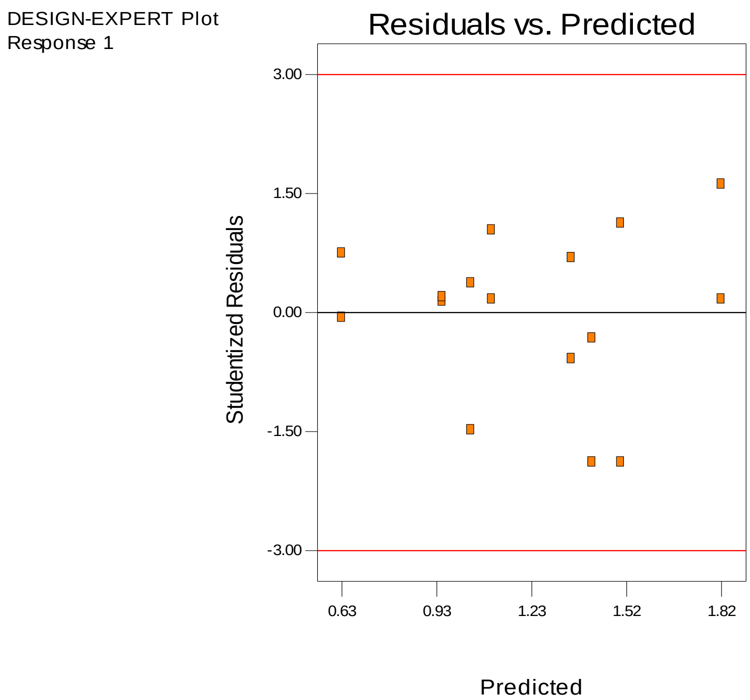
```
C*E        0.1350   0.0675
D*E       -0.3150  -0.1575
```



From the graph shown above, E, BE, and DE are important effects. From the effect estimation from table above these same effects are computed to be the largest effects (in absolute value).

(d) For the model with E, BE, and DE the normality assumption and constant variance seem to be reasonable.



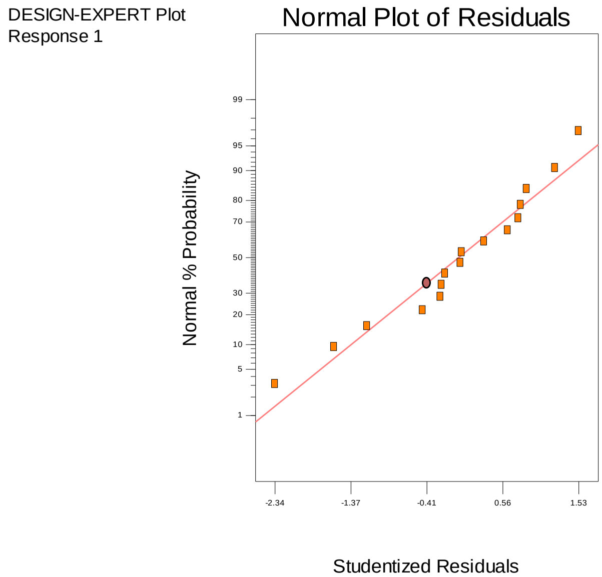


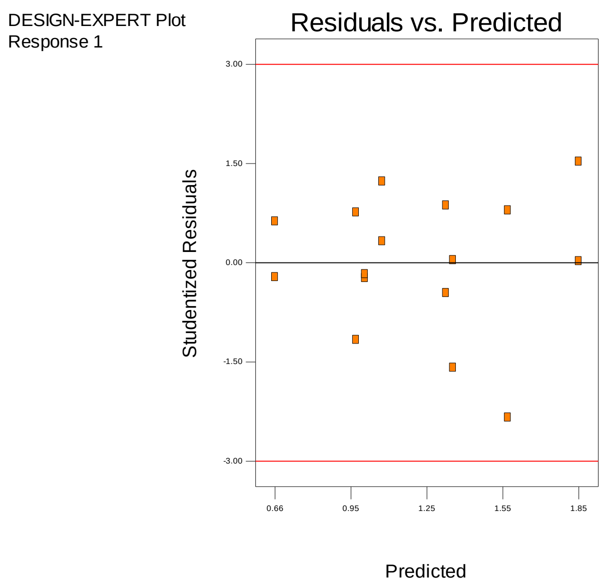
For a hierarchical model, the effects to include are B, D, E, BE, and DE. The ANOVA and residual plot follow:

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	1.97	5	0.39	8.94	0.0019
B	*0.031*	*1*	*0.031*	*0.69*	*0.4242*
D	*5.625E-003*	*1*	*5.625E-003*	*0.13*	*0.7284*
E	*0.88*	*1*	*0.88*	*20.03*	*0.0012*
BE	*0.66*	*1*	*0.66*	*14.87*	*0.0032*
DE	*0.40*	*1*	*0.40*	*9.00*	*0.0134*
Residual	0.44	10	0.044		

```
Cor Total           2.41            15
```





(e) For the non-hierarchical model $\hat{y} = 1.23 - 0.24E - 0.20BE - 0.16DE$
This equation could be used for further study of the process.

14-52 Montgomery (2012) described a 2^{4-1} fractional factorial design used to study four factors in a chemical process. The factors are A = temperature, B = pressure, C = concentration, and D = stirring rate, and the response is filtration rate. The design and the data are as follows:

Run	*A*	*B*	*C*	*D* = *ABC*	Treatment Combination	Filtration Rate
1	–	–	–	–	(1)	45
2	+	–	–	+	*ad*	100
3	–	+	–	+	*bd*	45
4	+	+	–	–	*ab*	65
5	–	–	+	+	*cd*	75
6	+	–	+	–	*ac*	60
7	–	+	+	–	*bc*	80
8	+	+	+	+	*abcd*	96

(a) Write down the alias relationships.
(b) Estimate the factor effects. Which factor effects appear large?
(c) Project this design into a full factorial in the three apparently important factors and provide a practical interpretation of the results.

(a)
```
Design Generators:  D = ABC
Alias Structure
I + ABCD
A + BCD
B + ACD
C + ABD
D + ABC
AB + CD
AC + BD
AD + BC
```

(b)
```
Term          Effect       Coef
Constant                 70.750
A             19.000      9.500
B              1.500      0.750
C             14.000      7.000
D             16.500      8.250
A*B           -1.000     -0.500
A*C          -18.500     -9.250
A*D           19.000      9.500
```

A, C, D, AC, and AD have large estimated effects.

(c)
```
Estimated Effects and Coefficients for rate
Term          Effect       Coef  StDev Coef        T       P
Constant                 70.750      0.6374   111.00   0.000
A             19.000      9.500      0.6374    14.90   0.004
C             14.000      7.000      0.6374    10.98   0.008
D             16.500      8.250      0.6374    12.94   0.006
A*C          -18.500     -9.250      0.6374   -14.51   0.005
A*D           19.000      9.500      0.6374    14.90   0.004

Analysis of Variance for rate
Source              DF     Seq SS     Adj SS     Adj MS      F       P
Main Effects         3    1658.50    1658.50    552.833 170.10   0.006
2-Way Interactions   2    1406.50    1406.50    703.250 216.38   0.005
Residual Error       2       6.50       6.50      3.250
Total                7    3071.50
```

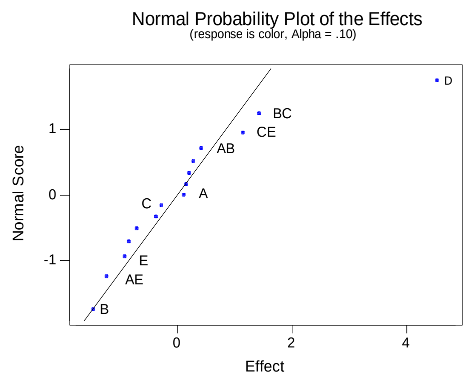
A, C, D, AC, and AD are significant. This appears to be an appropriate model for the data.

14-53 R. D. Snee ("Experimenting with a Large Number of Variables," in *Experiments in Industry: Design, Analysis and Interpretation of Results*, Snee, Hare, and Trout, eds., ASQC, 1985) described an experiment in which a 25–1 design with $I = ABCDE$ was used to investigate the effects of five factors on the color of a chemical product. The factors are A = solvent/reactant, B = catalyst/reactant, C = temperature, D = reactant purity, and E = reactant pH. The results obtained are as follows:

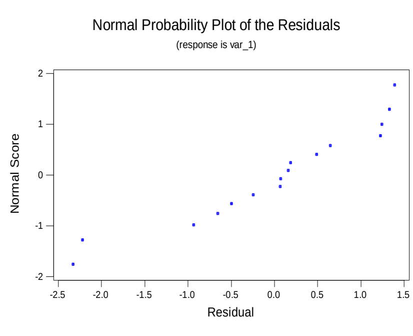
e	=	– 0.63	d	=	6.79
a	=	2.51	ade	=	6.47
b	=	– 2.68	bde	=	3.45
abe	=	1.66	abd	=	5.68
c	=	2.06	cde	=	5.22
ace	=	1.22	acd	=	4.38
bce	=	-2.09	bcd	=	4.30
abc	=	1.93	$abcde$	=	4.05

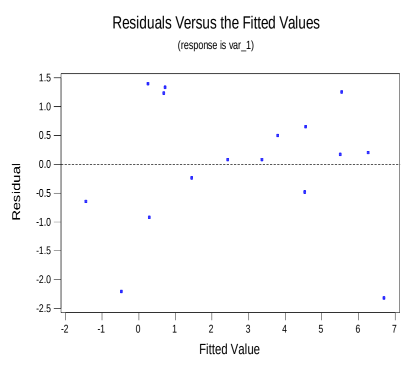
(a) Prepare a normal probability plot of the effects. Which factors are active?
(b) Calculate the residuals. Construct a normal probability plot of the residuals and plot the residuals versus the fitted values. Comment on the plots.
(c) If any factors are negligible, collapse the 2^{5-1} design into a full factorial in the active factors. Comment on the resulting design, and interpret the results.

(a) Several factors and interactions are potentially significant.

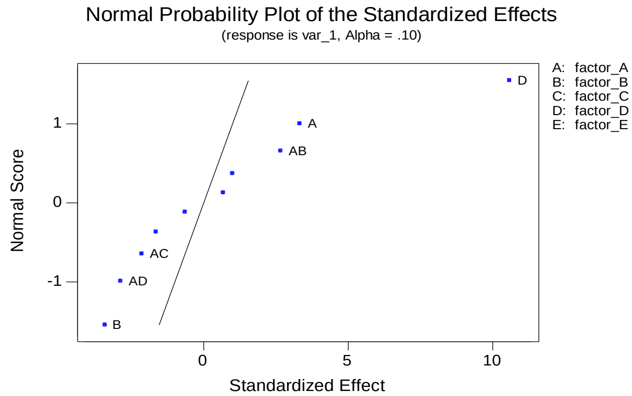


(b) There are no serious problems with the residual plots. The normal probability plot has some curvature and there is a little more variability at the lower and higher ends of the fitted values.





(c) Normal probability plot shows that we can collapse using only factors A, B, and D



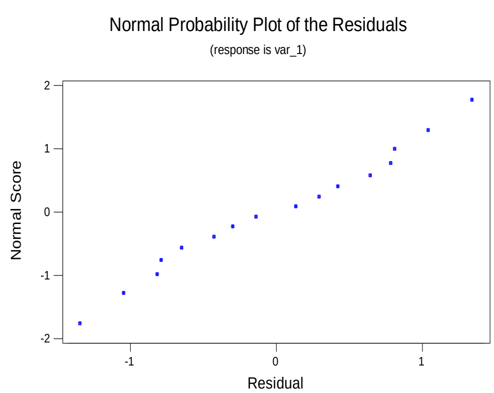
Estimated Effects and Coefficients for var_1

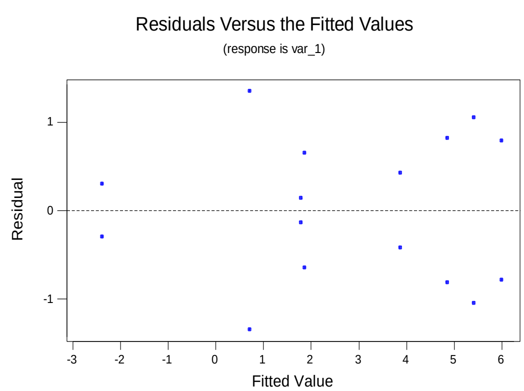
Term	Effect	Coef	StDev Coef	T	P
Constant		2.7700	0.2762	10.03	0.000
factor_A	1.4350	0.7175	0.2762	2.60	0.032
factor_B	-1.4650	-0.7325	0.2762	-2.65	0.029
factor_D	4.5450	2.2725	0.2762	8.23	0.000
factor_A*factor_B	1.1500	0.5750	0.2762	2.08	0.071
factor_A*factor_D	-1.2300	-0.6150	0.2762	-2.23	0.057
factor_B*factor_D	0.1200	0.0600	0.2762	0.22	0.833
factor_A*factor_B*factor_D	-0.3650	-0.1825	0.2762	-0.66	0.527

Analysis of Variance for var_1

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	99.450	99.4499	33.1500	27.15	0.000
2-Way Interactions	3	11.399	11.3992	3.7997	3.11	0.088
3-Way Interactions	1	0.533	0.5329	0.5329	0.44	0.527
Residual Error	8	9.767	9.7668	1.2208		
Pure Error	8	9.767	9.7668	1.2208		
Total	15	121.149				

Factors A, B, D, AB and AD are significant.





The normal probability plot does not indicate problesm. The reduced model ignores factor C and it is two replicates of a full factorial experiment in factors A, B, and D. There are 8 unique test points with two replicates at each. The model shown has 8 coefficients so that the fitted value is the mean of the replicates at each of the 8 unique test points. Therefore, at each unique test point there are equal positive and negative residuals. Consequently, the plot of residuals versus fitted values has symmetry about zero.

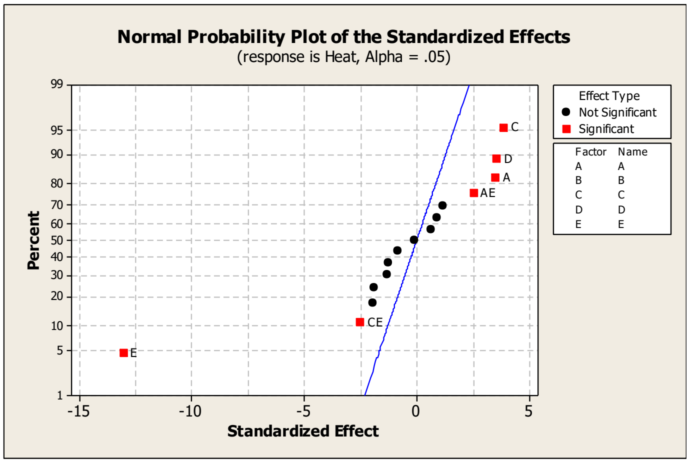
14-54 An article in *Quality Engineering* ["A Comparison of Multi-Response Optimization: Sensitivity to Parameter Selection" (1999, Vol. 11, pp. 405–415)] conducted a half replicate of a 2^5 factorial design to optimize the retort process of beef stew MREs, a military ration. The design factors are x_1 = sauce viscosity, x_2 = residual gas,

x_3 = solid/liquid ratio, x_4 = net weight, x_5 = rotation speed. The response variable is the heating rate index, a measure of heat penetration, and there are two replicates.

						Heating Rate Index	
Run	x_1	x_2	x_3	x_4	x_5	**I**	**II**
1	–1	–1	–1	–1	1	8.46	9.61
2	1	–1	–1	–1	–1	15.68	14.68
3	–1	1	–1	–1	–1	14.94	13.09
4	1	1	–1	–1	1	12.52	12.71
5	–1	–1	1	–1	–1	17.0	16.36
6	1	–1	1	–1	1	11.44	11.83
7	–1	1	1	–1	1	10.45	9.22
8	1	1	1	–1	–1	19.73	16.94
9	–1	–1	–1	1	–1	17.37	16.36
10	1	–1	–1	1	1	14.98	11.93
11	–1	1	–1	1	1	8.40	8.16
12	1	1	–1	1	–1	19.08	15.40
13	–1	–1	1	1	1	13.07	10.55
14	1	–1	1	1	–1	18.57	20.53
15	–1	1	1	1	–1	20.59	21.19
16	1	1	1	1	1	14.03	11.31

(a) Estimate the factor effects. Based on a normal probability plot of the effect estimates, identify a model for the data from this experiment.
(b) Conduct an ANOVA based on the model identified in part (a). What are your conclusions?
(c) Analyze the residuals and comment on model adequacy.
(d) Find a regression model to predict yield in terms of the coded factor levels.
(e) This experiment was replicated, so an ANOVA could have been conducted without using a normal plot of the Effects to tentatively identify a model. What model would be appropriate? Use the ANOVA to analyze this model and compare the results with those obtained from the normal probability plot approach.

(a)
```
Estimated Effects and Coefficients for Heat (coded units)

Term        Effect     Coef  SE Coef       T      P
Constant            14.256   0.2370   60.15  0.000
A            1.659   0.829   0.2370    3.50  0.003
B           -0.041  -0.021   0.2370   -0.09  0.932
C            1.840   0.920   0.2370    3.88  0.001
D            1.679   0.839   0.2370    3.54  0.003
E           -6.178  -3.089   0.2370  -13.03  0.000
A*B          0.301   0.151   0.2370    0.64  0.534
A*C         -0.915  -0.457   0.2370   -1.93  0.071
A*D         -0.391  -0.196   0.2370   -0.83  0.421
A*E          1.195   0.598   0.2370    2.52  0.023
B*C          0.555   0.278   0.2370    1.17  0.259
B*D         -0.609  -0.304   0.2370   -1.28  0.217
B*E         -0.593  -0.296   0.2370   -1.25  0.229
C*D          0.430   0.215   0.2370    0.91  0.378
C*E         -1.199  -0.599   0.2370   -2.53  0.022
D*E         -0.905  -0.453   0.2370   -1.91  0.074
```



The model is

$$\hat{y} = 14.2546 + 0.829x_1 + 0.920x_3 + 0.839x_4 - 3.809x_5 + 0.598x_1x_5 - 0.599x_3x_5$$

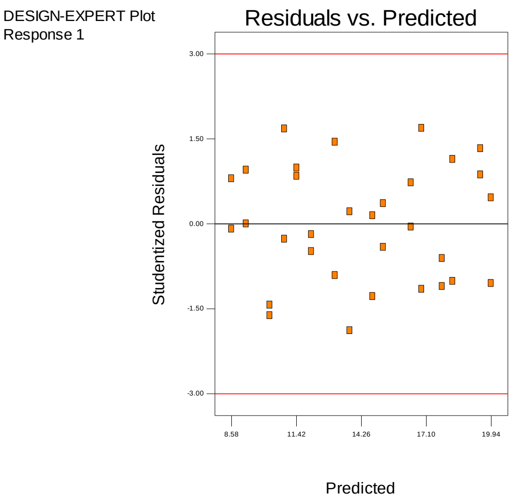
(b)

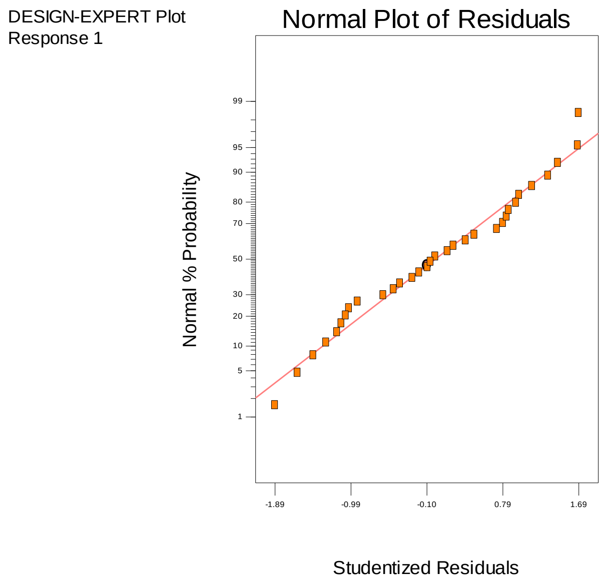
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	399.85	6	66.64	31.03	< 0.0001
A	*22.01*	*1*	*22.01*	*10.25*	*0.0037*
C	*27.08*	*1*	*27.08*	*12.61*	*0.0016*
D	*22.55*	*1*	*22.55*	*10.50*	*0.0034*
E	*305.29*	*1*	*305.29*	*142.16*	*< 0.0001*
AE	*11.42*	*1*	*11.42*	*5.32*	*0.0297*
CE	*11.50*	*1*	*11.50*	*5.35*	*0.0292*
Residual	53.69	25	2.15		
Lack of Fit	*24.93*	*9*	*2.77*	*1.54*	*0.2157*
Pure Error	*28.76*	*16*	*1.80*		
Cor Total	453.54	31			

The model is significant with significant main effects and two-factor interactions.

(c) The residual plots do not show any violations of the assumptions.





(d) The actual factor levels are not provided so only the model in the coded variables can be presented

$$\hat{y} = 14.2546 - 0.829x_1 + 0.920x_3 + 0.839x_4 - 3.809x_5 + 0.598x_1x_5 - 0.599x_3x_5$$

(e) Use the t-test to test individual effect as shown below

Term	Effect	Coef	SE Coef	T	P
Constant		14.256	0.2370	60.15	0.000
A	1.659	0.829	0.2370	3.50	0.003
B	-0.041	-0.021	0.2370	-0.09	0.932
C	1.840	0.920	0.2370	3.88	0.001
D	1.679	0.839	0.2370	3.54	0.003
E	-6.178	-3.089	0.2370	-13.03	0.000
A*B	0.301	0.151	0.2370	0.64	0.534
A*C	-0.915	-0.457	0.2370	-1.93	0.071
A*D	-0.391	-0.196	0.2370	-0.83	0.421
A*E	1.195	0.598	0.2370	2.52	0.023
B*C	0.555	0.278	0.2370	1.17	0.259
B*D	-0.609	-0.304	0.2370	-1.28	0.217
B*E	-0.593	-0.296	0.2370	-1.25	0.229
C*D	0.430	0.215	0.2370	0.91	0.378
C*E	-1.199	-0.599	0.2370	-2.53	0.022
D*E	-0.905	-0.453	0.2370	-1.91	0.074

At $\alpha = 0.05$, the t-test provides the same result as using normal probability plot in part (a).

14-55 An article in *Industrial and Engineering Chemistry* ["More on Planning Experiments to Increase Research Efficiency" (1970, pp. 60–65)] uses a 2^{5-2} design to investigate the effect on process yield of A = condensation temperature, B = amount of material 1, C = solvent volume, D = condensation time, and E = amount of material 2. The results obtained are as follows:

ae	=	23.2	*cd*	=	23.8
ab	=	15.5	*ace*	=	23.4
ad	=	16.9	*bde*	=	16.8
bc	=	16.2	*abcde*	=	18.1

(a) Verify that the design generators used were $I = ACE$ and $I = BDE$.
(b) Write down the complete defining relation and the aliases from the design.
(c) Estimate the main effects.

(d) Prepare an analysis of variance table. Verify that the *AB* and *AD* interactions are available to use as error.
(e) Plot the residuals versus the fitted values. Also construct a normal probability plot of the residuals. Comment on the results.

(a) The design generators are I = ACE and I = BDE. This is verified by looking at the following table.
The contrast for E is calculated using E = AC and the contrast for D is calculated using D = BE.

A	B	C	D	E	reponse
-1	-1	-1	-1	1	23.2
1	1	-1	-1	-1	15.5
1	-1	-1	1	-1	16.9
-1	1	1	-1	-1	16.2
-1	-1	1	1	-1	23.8
1	-1	1	-1	1	23.4
-1	1	-1	1	1	16.8
1	1	1	1	1	18.1

(b) Design Generator: D = BE, E = AC
Defining Relation: I = ACE = BDE =ABCDE

Aliases
A=CE=BCDE=ABDE
B=DE=ACDE=ABCE
C=AE=ABDE=BCDE
D=BE=ABCE=ACDE
E=AC=BD=ABCD

(c) Estimated Effects and Coefficients for response (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		19.238	0.7871	24.44	0.002
A	-1.525	-0.762	0.7871	-0.97	0.435
B	-5.175	-2.587	0.7871	-3.29	0.081
C	2.275	1.138	0.7871	1.45	0.285
D	-0.675	-0.337	0.7871	-0.43	0.710
E	2.275	1.137	0.7871	1.45	0.285

(d) Estimated Effects and Coefficients for response (coded units)

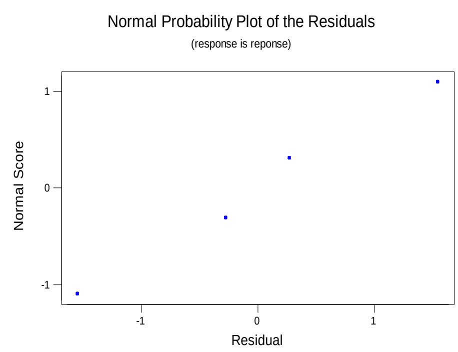
Term	Effect	Coef	SE Coef	T	P
Constant		19.238	1.138	16.91	0.038
A	-1.525	-0.762	1.138	-0.67	0.624
B	-5.175	-2.587	1.138	-2.27	0.264
C	2.275	1.138	1.138	1.00	0.500
D	-0.675	-0.337	1.138	-0.30	0.816
A*B	1.825	0.913	1.138	0.80	0.570
A*D	-1.275	-0.638	1.138	-0.56	0.675

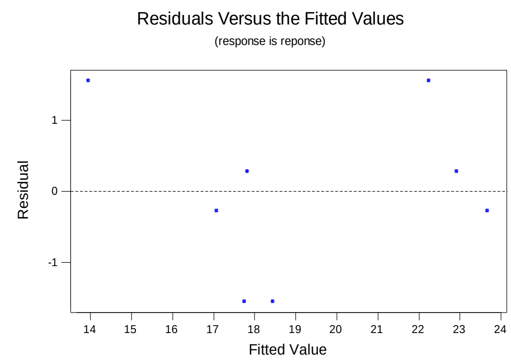
Analysis of Variance for response (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	69.475	69.475	17.369	1.68	0.517
2-Way Interactions	2	9.913	9.913	4.956	0.48	0.715
Residual Error	1	10.351	10.351	10.351		
Total	7	89.739				

Interactions AD and AB are not significant in the model, and therefore may be used as error.

(e) The normal probability plot and the plot of the residuals versus fitted values are satisfactory.



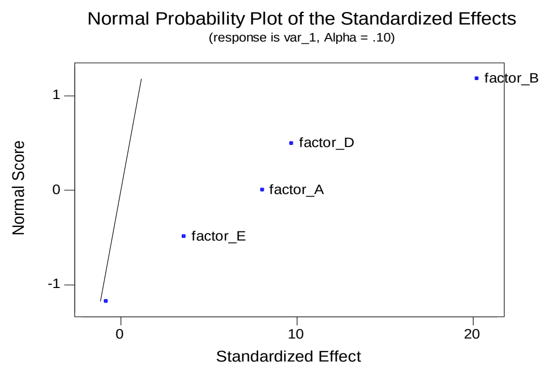


14-56 Suppose that in Exercise 14-16 only a $^1/_4$ fraction of the 2^5 design could be run. Construct the design and analyze the data that are obtained by selecting only the response for the eight runs in your design.

Generators D = AB, E = AC for 2^{5-2}, Resolution III

A	B	C	D	E	var_1
-1	-1	-1	1	1	1900
1	-1	-1	-1	-1	900
-1	1	-1	-1	1	3500
1	1	-1	1	-1	6100
-1	-1	1	1	-1	800
1	-1	1	-1	1	1200
-1	1	1	-1	-1	3000
1	1	1	1	1	6800

The normal probability plot and table below show that factors A, B, and D are significant.



Estimated Effects and Coefficients for var_1 (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		3025.00	90.14	33.56	0.001
factor_A	1450.00	725.00	90.14	8.04	0.015
factor_B	3650.00	1825.00	90.14	20.25	0.002
factor_C	-150.00	-75.00	90.14	-0.83	0.493
factor_D	1750.00	875.00	90.14	9.71	0.010
factor_E	650.00	325.00	90.14	3.61	0.069

Analysis of Variance for var_1 (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	37865000	37865000	7573000	116.51	0.009

```
Residual Error         2      130000     130000       65000
Total                  7    37995000
```

Factors A, B and D are significant. In these factors, the design is a 2^2 with two replicates.

14-57 For each of the following designs, write down the aliases, assuming that only main effects and two factor interactions are of interest.

(a) 2_{III}^{6-3} (b) 2_{IV}^{8-4}

(a) 2_{III}^{6-3}

```
Alias Structure

I + ABD + ACE + BCF + DEF + ABEF + ACDF + BCDE

A + BD + CE
B + AD + CF
C + AE + BF
D + AB + EF
E + AC + DF
F + BC + DE
AF + BE + CD
```

(b) 2_{IV}^{8-4}

```
Alias Structure

I + ABCG + ABDH + ABEF + ACDF + ACEH + ADEG + AFGH + BCDE + BCFH + BDFG + BEGH
+ CDGH + CEFG + DEFH

A
B
C
D
E
F
G
H
AB + CG + DH + EF
AC + BG + DF + EH
AD + BH + CF + EG
AE + BF + CH + DG
AF + BE + CD + GH
AG + BC + DE + FH
AH + BD + CE + FG
```

14-58 Consider the 2^{6-2} design in Table 14-29.

(a) Suppose that after analyzing the original data, we find that factors C and E can be dropped. What type of 2^k design is left in the remaining variables?
(b) Suppose that after the original data analysis, we find that factors D and F can be dropped. What type of 2^k design is left in the remaining variables? Compare the results with part (a). Can you explain why the answers are different?

(a) Because factors A, B, C, and E form a word in the complete defining relation, it can be verified that the resulting design is two replicates of a 2^{4-1} fractional factorial.

(b) Because factors A, B, C, and E form a word in the complete defining relation, it can be verified that the resulting design is two replicates of a 2^{4-1} fractional factorial. This is different than the design that results when C and E are dropped from the 2^{6-2}. When C and E are dropped, the result is a full factorial because the factors ABDF do not form a word in the complete defining relation.

14-59 An article in the *Journal of Radioanalytical and Nuclear Chemistry* (2008, Vol. 276(2), pp. 323–328) presented a 2^{8-4} fractional factorial design to identify sources of Pu contamination in the radioactivity material analysis of dried shellfish at the National Institute of Standards and Technology (NIST). The data are shown in the following table. No contamination occurred at runs 1, 4, and 9. The factors and levels are shown in the following table.

Factor	–1	+1
Glassware	Distilled water	Soap, acid, stored
Reagent	New	Old
Sample prep	Coprecipitation	Electrodeposition
Tracer	Stock	Fresh
Dissolution	Without	With
Hood	B	A
Chemistry	Without	With
Ashing	Without	With

2^{8-4}	Glassware	Reagent	Sample Prep	Tracer	Dissolution	Hood	Chemistry	Ashing	mBq
Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y
1	–1	–1	–1	–1	–1	–1	–1	–1	0
2	+1	–1	–1	–1	–1	+1	+1	+1	3.31
3	–1	+1	–1	–1	+1	–1	+1	+1	0.0373
4	+1	+1	–1	–1	+1	+1	–1	–1	0
5	–1	–1	+1	–1	+1	+1	+1	–1	0.0649
6	+1	–1	+1	–1	+1	–1	–1	+1	0.133
7	–1	+1	+1	–1	–1	+1	–1	+1	0.0461
8	+1	+1	+1	–1	–1	–1	+1	–1	0.0297
9	–1	–1	–1	+1	+1	+1	–1	+1	0
10	+1	–1	–1	+1	+1	–1	+1	–1	0.287
11	–1	+1	–1	+1	–1	+1	+1	–1	0.133
12	+1	+1	–1	+1	–1	–1	–1	+1	0.0476
13	–1	–1	+1	+1	–1	–1	+1	+1	0.133
14	+1	–1	+1	+1	–1	+1	–1	–1	5.75
15	–1	+1	+1	+1	+1	–1	–1	–1	0.0153
16	+1	+1	+1	+1	+1	+1	+1	+1	2.47

(a) Write down the alias relationships.
(b) Estimate the main effects.
(c) Prepare a normal probability plot for the effects and interpret the results.

(a) Suppose A = x_1, B = x_2, C = x_3, D = x_4, E = x_5, F = x_6, G = x_7, H = x_8.

Generators are computer software defaults.

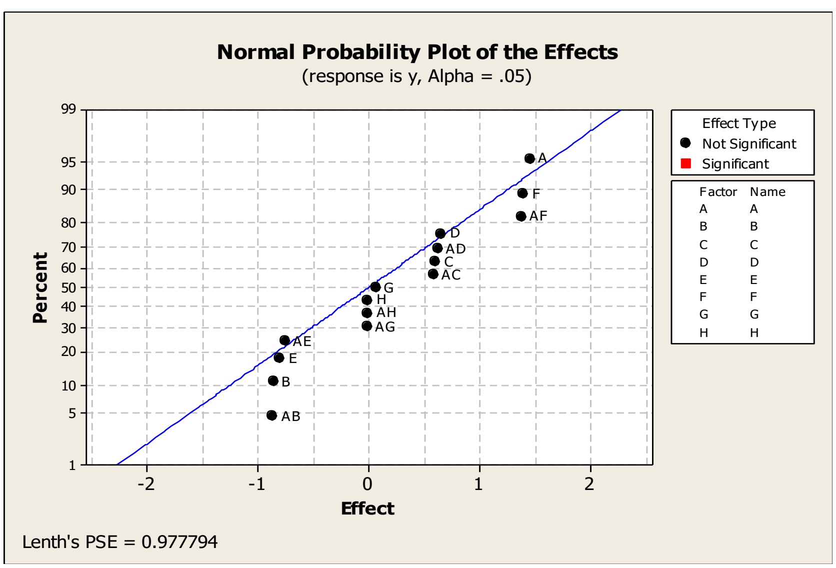
```
Design Generators: E = BCD, F = ACD, G = ABC, H = ABD


Alias Structure (up to order 4)

I + ABCG + ABDH + ABEF + ACDF + ACEH + ADEG + AFGH + BCDE + BCFH + BDFG + BEGH
     + CDGH + CEFG + DEFH

A + BCG + BDH + BEF + CDF + CEH + DEG + FGH
```

```
B + ACG + ADH + AEF + CDE + CFH + DFG + EGH
C + ABG + ADF + AEH + BDE + BFH + DGH + EFG
D + ABH + ACF + AEG + BCE + BFG + CGH + EFH
E + ABF + ACH + ADG + BCD + BGH + CFG + DFH
F + ABE + ACD + AGH + BCH + BDG + CEG + DEH
G + ABC + ADE + AFH + BDF + BEH + CDH + CEF
H + ABD + ACE + AFG + BCF + BEG + CDG + DEF
AB + CG + DH + EF + ACDE + ACFH + ADFG + AEGH + BCDF + BCEH + BDEG + BFGH
AC + BG + DF + EH + ABDE + ABFH + ADGH + AEFG + BCDH + BCEF + CDEG + CFGH
AD + BH + CF + EG + ABCE + ABFG + ACGH + AEFH + BCDG + BDEF + CDEH + DFGH
AE + BF + CH + DG + ABCD + ABGH + ACFG + ADFH + BCEG + BDEH + CDEF + EFGH
AF + BE + CD + GH + ABCH + ABDG + ACEG + ADEH + BCFG + BDFH + CEFH + DEFG
AG + BC + DE + FH + ABDF + ABEH + ACDH + ACEF + BDGH + BEFG + CDFG + CEGH
AH + BD + CE + FG + ABCF + ABEG + ACDG + ADEF + BCGH + BEFH + CDFH + DEGH
```

(b)

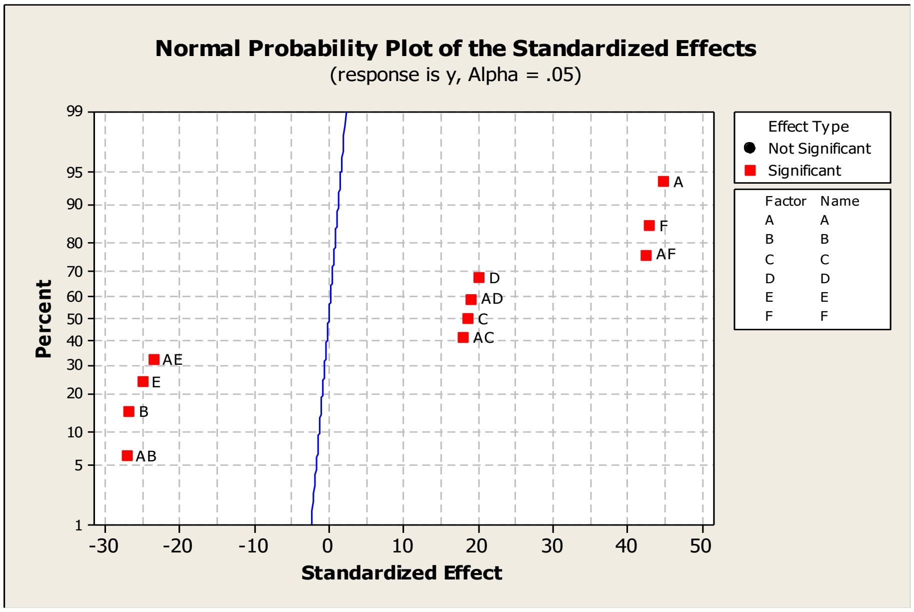
```
Estimated Effects and Coefficients for y (coded units)

Term        Effect     Coef
Constant             0.7786
A           1.4497   0.7249
B          -0.8624  -0.4312
C           0.6034   0.3017
D           0.6519   0.3259
E          -0.8052  -0.4026
F           1.3864   0.6932
G           0.0591   0.0296
H          -0.0129  -0.0064
A*B        -0.8708  -0.4354
A*C         0.5811   0.2906
A*D         0.6186   0.3093
A*E        -0.7566  -0.3783
A*F         1.3718   0.6859
A*G        -0.0176  -0.0088
A*H        -0.0137  -0.0068
```

(c) The normal probability plot of the effects follows.



From the effects table and the normal probability plot effects G, H, AG, and AH are smaller than the others. If these are used to estimate error the following estimates and normal plot are obtained.

```
Estimated Effects and Coefficients for y (coded units)

Term        Effect     Coef  SE Coef       T      P
Constant             0.7786  0.01612   48.31  0.000
A           1.4497   0.7249  0.01612   44.98  0.000
B          -0.8624  -0.4312  0.01612  -26.75  0.000
C           0.6034   0.3017  0.01612   18.72  0.000
D           0.6519   0.3259  0.01612   20.22  0.000
E          -0.8052  -0.4026  0.01612  -24.98  0.000
F           1.3864   0.6932  0.01612   43.01  0.000
A*B        -0.8708  -0.4354  0.01612  -27.02  0.000
A*C         0.5811   0.2906  0.01612   18.03  0.000
A*D         0.6186   0.3093  0.01612   19.19  0.000
A*E        -0.7566  -0.3783  0.01612  -23.47  0.000
A*F         1.3718   0.6859  0.01612   42.56  0.000


S = 0.0644648   R-Sq = 99.96%   R-Sq(adj) = 99.85%

Analysis of Variance for y (coded units)

Source              DF   Seq SS   Adj SS   Adj MS       F      P
Main Effects         6  24.8193  24.8193  4.13654  995.39  0.000
2-Way Interactions   5  15.7318  15.7318  3.14635  757.11  0.000
Residual Error       4   0.0166   0.0166  0.00416
Total               15  40.5676
```



With this estimate of error, the remaining effects are all significant. Also, any interpretations of these effects need to consider the aliases from the alias structure shown previously.

14-60 An article in the *Journal of Marketing Research* (1973, Vol. 10(3), pp. 270–276) presented a 2^{7-4} fractional factorial design to conduct marketing research:

Runs	*A*	*B*	*C*	*D*	*E*	*F*	*G*	Sales for a 6-Week Period (in $1000)
1	–1	–1	–1	1	1	1	–1	8.7
2	1	–1	–1	–1	–1	1	1	15.7
3	–1	1	–1	–1	1	–1	1	9.7
4	1	1	–1	1	–1	–1	–1	11.3
5	–1	–1	1	1	–1	–1	1	14.7
6	1	–1	1	–1	1	–1	–1	22.3
7	–1	1	1	–1	–1	1	–1	16.1
8	1	1	1	1	1	1	1	22.1

The factors and levels are shown in the following table.

	Factor	–1	+1
A	Television advertising	No advertising	Advertising
B	Billboard advertising	No advertising	Advertising
C	Newspaper advertising	No advertising	Advertising
D	Candy wrapper design	Conservative design	Flashy design
E	Display design	Normal shelf display	Special aisle display
F	Free sample program	No free samples	Free samples
G	Size of candy bar	1 oz bar	2½ oz bar

(a) Write down the alias relationships.
(b) Estimate the main effects.
(c) Prepare a normal probability plot for the effects and interpret the results.

(a)

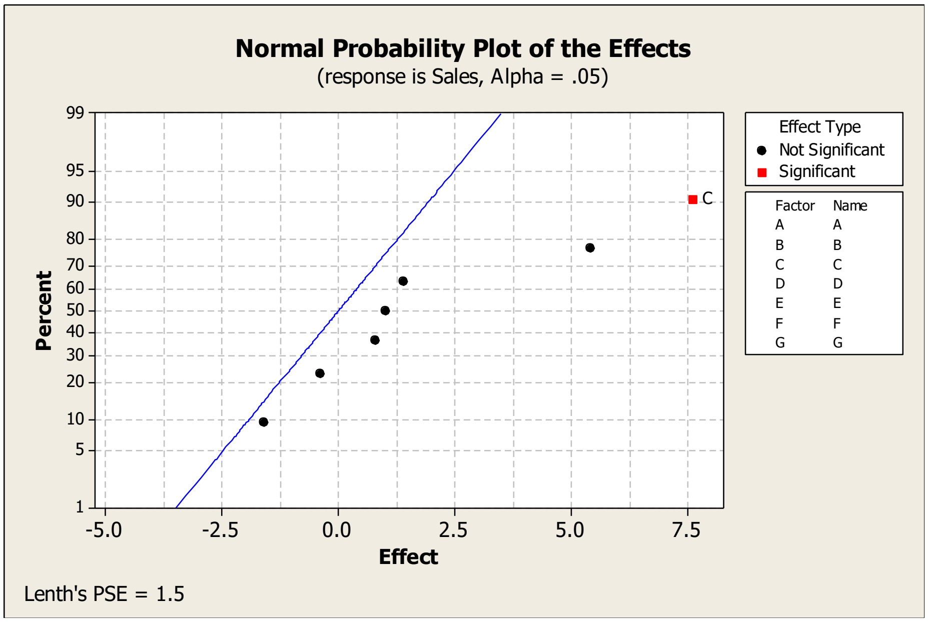
```
Alias Structure (up to order 3)
I + A*B*D + A*C*E + A*F*G + B*C*F + B*E*G + C*D*G + D*E*F
A + B*D + C*E + F*G + B*C*G + B*E*F + C*D*F + D*E*G
B + A*D + C*F + E*G + A*C*G + A*E*F + C*D*E + D*F*G
C + A*E + B*F + D*G + A*B*G + A*D*F + B*D*E + E*F*G
D + A*B + C*G + E*F + A*C*F + A*E*G + B*C*E + B*F*G
E + A*C + B*G + D*F + A*B*F + A*D*G + B*C*D + C*F*G
F + A*G + B*C + D*E + A*B*E + A*C*D + B*D*G + C*E*G
G + A*F + B*E + C*D + A*B*C + A*D*E + B*D*F + C*E*F
```

(b)
Factorial Fit: Sales versus A, B, C, D, E, F, G

```
Estimated Effects and Coefficients for Sales (coded units)

Term        Effect     Coef
Constant             15.0000
A           5.4000    2.7000
B          -0.4000   -0.2000
```

```
C           7.6000   3.8000
D          -1.6000  -0.8000
E           1.4000   0.7000
F           1.0000   0.5000
G           0.8000   0.4000
```

(c) The plot indicates that only Factor C is a significant effect, but one might also consider the effect of A as sufficiently distant from the line to be considered significant..



14-61 An article in *Bioresource Technology* ["Medium Optimization for Phenazine-1-carboxylic Acid Production by a gacA qscR Double Mutant of Pseudomonas sp. M18 Using Response Surface Methodology" (Vol. 101(11), 2010, pp. 4089-4095)] described an experiment to optimize culture medium factors to enhance phenazine-1-carboxylic acid (PCA) production. A 2^{5-1} fractional factorial design was conducted with factors soybean meal, glucose, corn steep liquor, ethanol, and $MgSO_4$. Rows below the horizontal line in the table (coded with zeros) correspond to center points.

(a) What is the generator of this design?
(b) What is the resolution of this design?
(c) Analyze factor effects and comment on important ones.
(d) Develop a regression model to predict production in terms of the actual factor levels.
(e) Does a residual analysis indicate any problems?

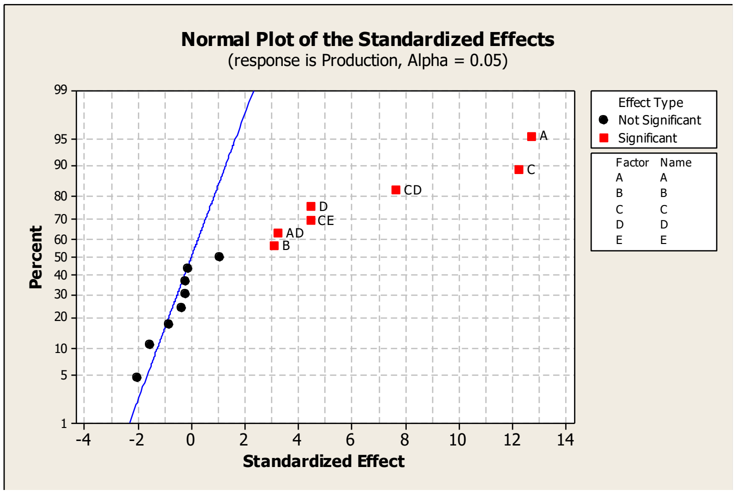
Run	X_1	X_2	X_3	X_4	X_5	Production (g/L)
1	−	−	−	−	+	1575.5
2	+	−	−	−	−	2201.4
3	−	+	−	−	−	1813.9
4	+	+	−	−	+	2164.1
5	−	−	+	−	−	1739.6
6	+	−	+	−	+	2483.2
7	−	+	+	−	+	2159.1
8	+	+	+	−	−	2257.7
9	−	−	−	+	−	1386.3
10	+	−	−	+	+	1967.8
11	−	+	−	+	+	1306.0
12	+	+	−	+	−	2486.9
13	−	−	+	+	+	2374.9
14	+	−	+	+	−	2932.7
15	−	+	+	+	−	2458.9
16	+	+	+	+	+	3204.9
17	0	0	0	0	0	2630.4
18	0	0	0	0	0	2571.6
19	0	0	0	0	0	2734.5
20	0	0	0	0	0	2480.4
21	0	0	0	0	0	2662.5

Variable	Component	Levels (g/L)		
		−1	0	+1
X_1	Soybean meal	30	45	60
X_2	Ethanol	12	18	24
X_3	Corn steep liquor	7	11	14
X_4	Glucose	10	15	20
X_5	$MgSO_4$	0	1	2

(a) Generator: E = ABCD

(b) Resolution = 5

(c) Estimated factor effects:




```
Alias Structure

I + ABCDE

A + BCDE
B + ACDE
C + ABDE
D + ABCE
E + ABCD
AB + CDE
AC + BDE
AD + BCE
AE + BCD
BC + ADE
BD + ACE
BE + ACD
CD + ABE
CE + ABD
DE + ABC
```

Factorial Fit: Production versus A, B, C, D, E

Estimated Effects and Coefficients for Production (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		2157.06	23.97	89.99	0.000
A	610.56	305.28	23.97	12.74	0.000
B	148.76	74.38	23.97	3.10	0.036
C	588.64	294.32	23.97	12.28	0.000
D	215.49	107.74	23.97	4.50	0.011
E	-5.24	-2.62	23.97	-0.11	0.918
A*B	-16.64	-8.32	23.97	-0.35	0.746
A*C	-74.06	-37.03	23.97	-1.54	0.197
A*D	155.99	77.99	23.97	3.25	0.031
A*E	-9.44	-4.72	23.97	-0.20	0.854
B*C	-11.21	-5.61	23.97	-0.23	0.827
B*D	49.99	24.99	23.97	1.04	0.356
B*E	-40.59	-20.29	23.97	-0.85	0.445
C*D	367.46	183.73	23.97	7.67	0.002
C*E	213.54	106.77	23.97	4.45	0.011
D*E	-97.56	-48.78	23.97	-2.04	0.112
Ct Pt		458.82	49.12	9.34	0.001

```
S = 95.8782     PRESS = *
R-Sq = 99.25%   R-Sq(pred) = *%   R-Sq(adj) = 96.24%

Analysis of Variance for Production (coded units)

Source              DF   Seq SS   Adj SS   Adj MS       F      P
Main Effects         5  3151493  3151493   630299   68.57  0.001
  A                  1  1491146  1491146  1491146  162.21  0.000
  B                  1    88521    88521    88521    9.63  0.036
  C                  1  1385976  1385976  1385976  150.77  0.000
  D                  1   185739   185739   185739   20.21  0.011
  E                  1      110      110      110    0.01  0.918
2-Way Interactions  10   898402   898402    89840    9.77  0.021
  A*B                1     1107     1107     1107    0.12  0.746
  A*C                1    21941    21941    21941    2.39  0.197
  A*D                1    97328    97328    97328   10.59  0.031
  A*E                1      356      356      356    0.04  0.854
  B*C                1      503      503      503    0.05  0.827
  B*D                1     9995     9995     9995    1.09  0.356
  B*E                1     6589     6589     6589    0.72  0.445
  C*D                1   540115   540115   540115   58.76  0.002
  C*E                1   182393   182393   182393   19.84  0.011
  D*E                1    38074    38074    38074    4.14  0.112
  Curvature          1   801978   801978   801978   87.24  0.001
Residual Error       4    36771    36771     9193
  Pure Error         4    36771    36771     9193
Total               20  4888643
```

Main effects A B C D and two-factor interaction AD CD CE are significant. The higher-order interaction in each alias pair is ignored and it is assumed that the main effect or the second-order interaction is responsible for the effect. Curvature is also a significant effect.

(d) Build a model using only significant effects. Because CE is a significant interaction, to maintain a hierarchical model, factor E (MgSO4) is also added to the model.

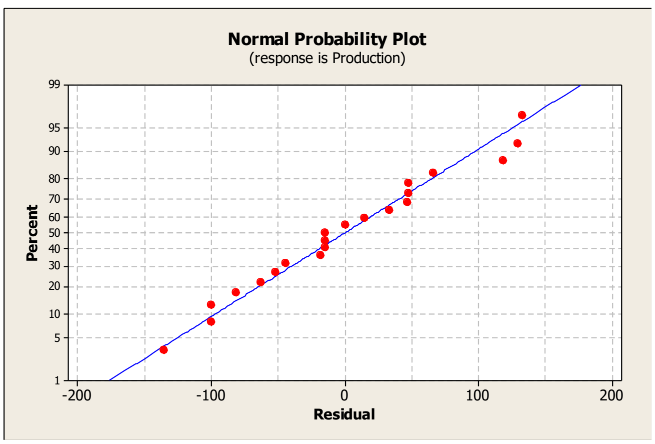
```
Estimated Effects and Coefficients for Production (coded units)

Term                          Effect     Coef  SE Coef      T      P
Constant                              2157.06    25.60  84.26  0.000
Soybean meal                  610.56   305.28    25.60  11.93  0.000
Ethanol                       148.76    74.38    25.60   2.91  0.014
Corn steep liquor             588.64   294.32    25.60  11.50  0.000
Glucose                       215.49   107.74    25.60   4.21  0.001
MgSO4                          -5.24    -2.62    25.60  -0.10  0.920
Soybean meal*Glucose          155.99    77.99    25.60   3.05  0.011
Corn steep liquor*Glucose     367.46   183.73    25.60   7.18  0.000
Corn steep liquor*MgSO4       213.54   106.77    25.60   4.17  0.002
Ct Pt                                  416.78    52.59   7.93  0.000

S = 102.397     PRESS = 467919
R-Sq = 97.64%   R-Sq(pred) = 90.43%   R-Sq(adj) = 95.71%

Analysis of Variance for Production (coded units)

Source                        DF   Seq SS   Adj SS   Adj MS       F      P
Main Effects                   5  3294941  3151493   630299   60.11  0.000
  Soybean meal                 1  1491146  1491146  1491146  142.22  0.000
  Ethanol                      1    88521    88521    88521    8.44  0.014
  Corn steep liquor            1  1529424  1385976  1385976  132.19  0.000
  Glucose                      1   185739   185739   185739   17.71  0.001
  MgSO4                        1      110      110      110    0.01  0.920
2-Way Interactions             3   819836   819836   273279   26.06  0.000
```

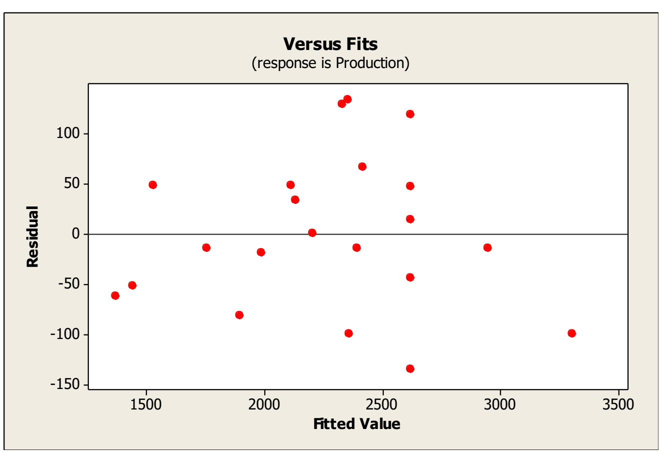
```
  Soybean meal*Glucose          1    97328    97328    97328    9.28  0.011
  Corn steep liquor*Glucose     1   540115   540115   540115   51.51  0.000
  Corn steep liquor*MgSO4       1   182393   182393   182393   17.40  0.002
  Curvature                     1   658530   658530   658530   62.81  0.000
Residual Error                 11   115336   115336    10485
  Lack of Fit                   7    78566    78566    11224    1.22  0.448
  Pure Error                    4    36771    36771     9193
Total                          20  4888643
```

Coefficients used to predict production in terms of the actual factor levels. Because curvature is significant, a second order model is needed here.

```
Estimated Coefficients for Production using data in uncoded units
Term                            Coef
Constant                     2490.33
Soybean meal                 4.75333
Ethanol                      12.3969
Corn steep liquor           -103.898
Glucose                     -135.486
MgSO4                       -322.925
Soybean meal*Glucose         1.03992
Corn steep liquor*Glucose    10.4989
Corn steep liquor*MgSO4      30.5054
Ct Pt                        416.778
```

(e) The residuals from the reduced model are plotted here. The plots do not indicate departures from assumptions.





14-62 An article in *Journal of Hazardous Materials* ["Statistical Factor-Screening and Optimization in Slurry Phase Bioremediation of 2,4,6-trinitrotoluene Contaminated Soil," (2011, Vol. 188(1), pp. 1–9)] described an experiment to optimize the removal of TNT 2,4,6-trinitrotoluene (TNT). TNT is a predominant contaminant at ammunition plants, testing facilities and military zones. TNT removal (TR) is measured by the percentage of the initial concentration removed (mg/kg-soil).
A 2^{7-3} fractional factorial design was conducted. The data are in the following table. Rows below the horizontal line in the table (coded with zeros) correspond to center points.

(a) What is the alias structure of this design?
(b) What is the resolution of this design?
(c) Analyze factor effects and comment on important effects.
(d) Develop a regression model to predict removal in terms of the actual factor levels.
(e) Does a residual analysis indicate any problems?

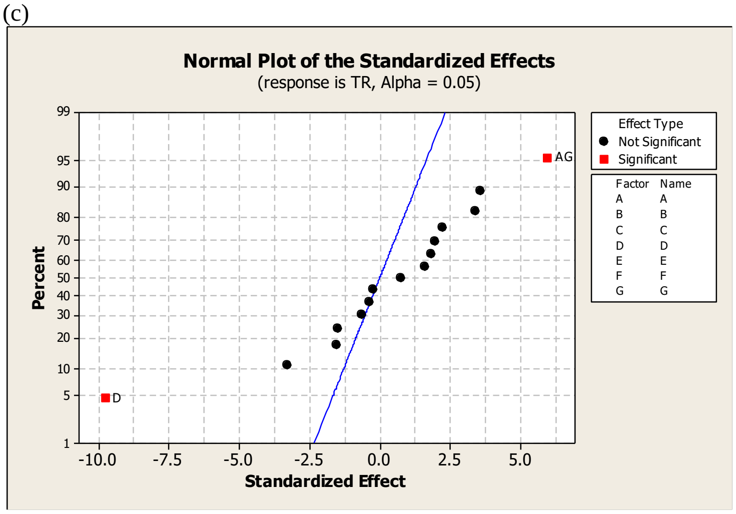
Run	Glucose (g/L) A	NH_4Cl (g/L) B	Tween80 (g/L) C	Slurry (g/ml) D	Temp (°C) E	Yeast (g/L) F	Inoculum (vol.%) G	TR
1	2	0.1	5	20	35	0.2	10	90.5
2	8	0.1	5	20	20	0.2	5	80.1
3	8	0.1	1	20	35	0	10	92.3
4	2	0.1	5	40	35	0	5	82.9
5	2	0.1	1	40	20	0.2	10	68.1
6	8	0.5	1	20	20	0.2	10	90.4
7	2	0.5	1	40	35	0	10	71.6
8	8	0.1	1	40	35	0.2	5	79.5
9	8	0.5	5	40	35	0.2	10	86.5
10	2	0.5	5	40	20	0.2	5	84.1
11	8	0.5	5	20	35	0	5	91.3
12	2	0.5	1	20	35	0.2	5	89.7
13	8	0.5	1	40	20	0	5	78.1
14	2	0.1	1	20	20	0	5	90.4
15	2	0.5	5	20	20	0	10	91
16	8	0.1	5	40	20	0	10	83.6
17	5	0.3	3	30	27.5	0.1	7.5	85.6
18	5	0.3	3	30	27.5	0.1	7.5	89.7
19	5	0.3	3	30	27.5	0.1	7.5	88.3

(a) E =ABC F=BCD G=ACD

```
Factors:    7   Base Design:         7, 16   Resolution:   IV
Runs:      19   Replicates:              1   Fraction:    1/8
Blocks:     1   Center pts (total):      3

Design Generators: E = ABC, F = BCD, G = ACD

Alias Structure

I + ABCE + ABFG + ACDG + ADEF + BCDF + BDEG + CEFG

A + BCE + BFG + CDG + DEF + ABCDF + ABDEG + ACEFG
B + ACE + AFG + CDF + DEG + ABCDG + ABDEF + BCEFG
C + ABE + ADG + BDF + EFG + ABCFG + ACDEF + BCDEG
D + ACG + AEF + BCF + BEG + ABCDE + ABDFG + CDEFG
E + ABC + ADF + BDG + CFG + ABEFG + ACDEG + BCDEF
F + ABG + ADE + BCD + CEG + ABCEF + ACDFG + BDEFG
```

```
G + ABF + ACD + BDE + CEF + ABCEG + ADEFG + BCDFG
AB + CE + FG + ACDF + ADEG + BCDG + BDEF + ABCEFG
AC + BE + DG + ABDF + AEFG + BCFG + CDEF + ABCDEG
AD + CG + EF + ABCF + ABEG + BCDE + BDFG + ACDEFG
AE + BC + DF + ABDG + ACFG + BEFG + CDEG + ABCDEF
AF + BG + DE + ABCD + ACEG + BCEF + CDFG + ABDEFG
AG + BF + CD + ABDE + ACEF + BCEG + DEFG + ABCDFG
BD + CF + EG + ABCG + ABEF + ACDE + ADFG + BCDEFG
ABD + ACF + AEG + BCG + BEF + CDE + DFG + ABCDEFG
```

(b) Resolution = 4
(c)




```
Estimated Effects and Coefficients for TR (coded units)

Term                    Effect     Coef  SE Coef       T      P
Constant                         84.381   0.5210  161.96  0.000
Glucose                  1.687    0.844   0.5210    1.62  0.247
NH4Cl                    1.912    0.956   0.5210    1.84  0.208
Tween80                  3.737    1.869   0.5210    3.59  0.070
Slurry                 -10.163   -5.081   0.5210   -9.75  0.010
Temp                     2.313    1.156   0.5210    2.22  0.157
Yeast                   -1.538   -0.769   0.5210   -1.48  0.278
Inoculum                -0.262   -0.131   0.5210   -0.25  0.825
Glucose*NH4Cl            0.788    0.394   0.5210    0.76  0.529
Glucose*Tween80         -3.438   -1.719   0.5210   -3.30  0.081
Glucose*Slurry           3.563    1.781   0.5210    3.42  0.076
Glucose*Temp             2.037    1.019   0.5210    1.96  0.190
Glucose*Yeast           -0.662   -0.331   0.5210   -0.64  0.590
Glucose*Inoculum         6.213    3.106   0.5210    5.96  0.027
NH4Cl*Slurry            -0.362   -0.181   0.5210   -0.35  0.761
Glucose*NH4Cl*Slurry    -1.588   -0.794   0.5210   -1.52  0.267
Ct Pt                             3.485   1.3112    2.66  0.117

S = 2.08407     PRESS = *
R-Sq = 98.98%   R-Sq(pred) = *%    R-Sq(adj) = 90.80%

Analysis of Variance for TR (coded units)

Source                  DF   Seq SS   Adj SS   Adj MS      F      P
Main Effects             7  526.124  526.124   75.161  17.30  0.056
  Glucose                1   11.391   11.391   11.391   2.62  0.247
  NH4Cl                  1   14.631   14.631   14.631   3.37  0.208
  Tween80                1   55.876   55.876   55.876  12.86  0.070
  Slurry                 1  413.106  413.106  413.106  95.11  0.010
```

```
  Temp                 1   21.391   21.391   21.391   4.92  0.157
  Yeast                1    9.456    9.456    9.456   2.18  0.278
  Inoculum             1    0.276    0.276    0.276   0.06  0.825
2-Way Interactions     7  273.779  273.779   39.111   9.00  0.104
  Glucose*NH4Cl        1    2.481    2.481    2.481   0.57  0.529
  Glucose*Tween80      1   47.266   47.266   47.266  10.88  0.081
  Glucose*Slurry       1   50.766   50.766   50.766  11.69  0.076
  Glucose*Temp         1   16.606   16.606   16.606   3.82  0.190
  Glucose*Yeast        1    1.756    1.756    1.756   0.40  0.590
  Glucose*Inoculum     1  154.381  154.381  154.381  35.54  0.027
  NH4Cl*Slurry         1    0.526    0.526    0.526   0.12  0.761
3-Way Interactions     1   10.081   10.081   10.081   2.32  0.267
  Glucose*NH4Cl*Slurry 1   10.081   10.081   10.081   2.32  0.267
  Curvature            1   30.690   30.690   30.690   7.07  0.117
Residual Error         2    8.687    8.687    4.343
  Pure Error           2    8.687    8.687    4.343
Total                 18  849.361
```

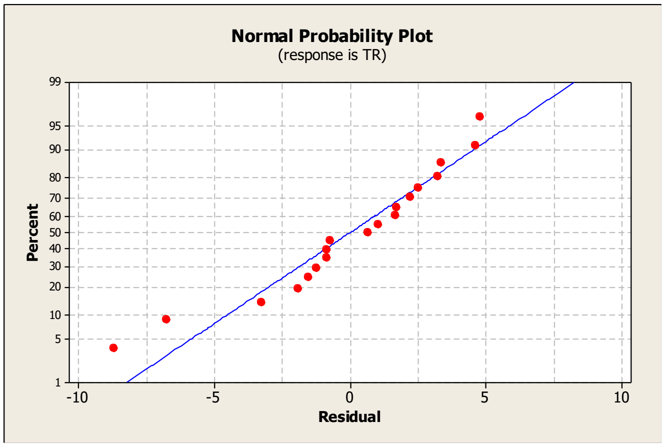
From the normal plot of the effects, we choose effects D and AG = BF = CD. Because D is significant, we attribute the interaction effect to CD. To maintain a hierarchical model we also select factor C. We also omit the curvature term in the model because it is not significant.

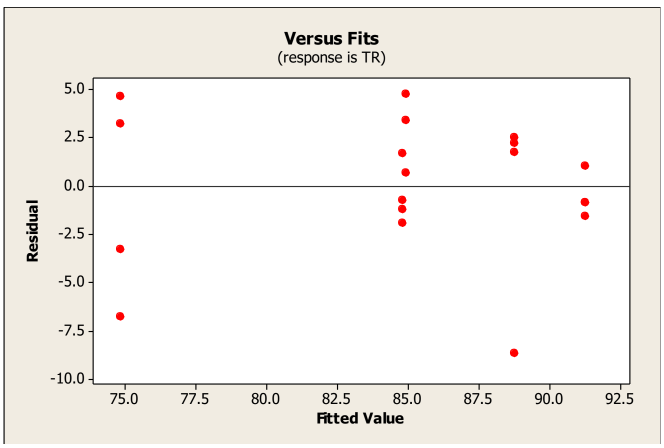
```
Estimated Effects and Coefficients for TR (coded units)

Term       Effect    Coef  SE Coef      T      P
Constant           84.932   0.8905  95.38  0.000
C           3.737   1.869   0.9704   1.93  0.073
D         -10.162  -5.081   0.9704  -5.24  0.000
C*D         6.212   3.106   0.9704   3.20  0.006


S = 3.88157      PRESS = 370.078
R-Sq = 73.39%   R-Sq(pred) = 56.43%   R-Sq(adj) = 68.07%

Analysis of Variance for TR (coded units)
Source              DF   Seq SS   Adj SS  Adj MS      F      P
Main Effects         2  468.981  468.981  234.49  15.56  0.000
  C                  1   55.876   55.876   55.88   3.71  0.073
  D                  1  413.106  413.106  413.11  27.42  0.000
2-Way Interactions   1  154.381  154.381  154.38  10.25  0.006
  C*D                1  154.381  154.381  154.38  10.25  0.006
Residual Error      15  225.999  225.999   15.07
  Curvature          1   30.690   30.690   30.69   2.20  0.160
  Pure Error        14  195.309  195.309   13.95
Total               18  849.361
```

(d) Coefficients in terms of the actual factor levels:

```
Estimated Coefficients for TR using data in uncoded units

Term                  Coef
Constant           111.350
Tween80           -3.72500
Slurry           -0.974062
Tween80*Slurry    0.155312
```

)e) Residuals plots do not indicate any violations of assumptions. There is a slight, but not serious, curvature in the plot of the residuals versus the fitted values.





Section 14-8

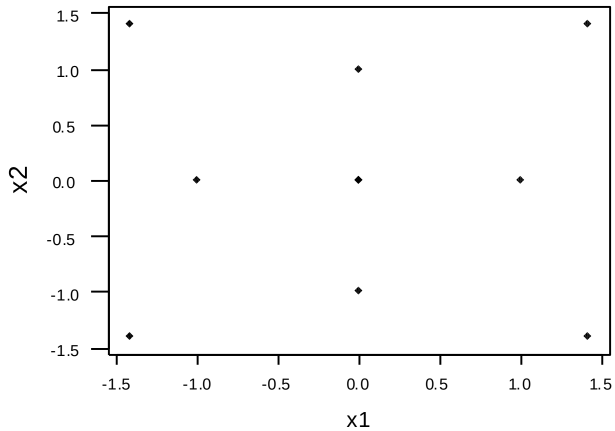
14-63 An article in *Rubber Age* (1961, Vol. 89, pp. 453–458) describes an experiment on the manufacture of a product in which two factors were varied. The factors are reaction time (hr) and temperature (°C). These factors are coded as x_1 = (time – 12) / 8 and x_2 = (temperature – 250) / 30. The following data were observed where y is the yield (in percent):

Run Number	x_1	x_2	y
1	–1	0	83.8
2	1	0	81.7
3	0	0	82.4
4	0	0	82.9
5	0	–1	84.7
6	0	1	75.9
7	0	0	81.2
8	–1.414	–1.414	81.3
9	–1.414	1.414	83.1
10	1.414	–1.414	85.3
11	1.414	1.414	72.7
12	0	0	82.0

(a) Plot the points at which the experimental runs were made.
(b) Fit a second-order model to the data. Is the second-order model adequate?

(c) Plot the yield response surface. What recommendations would you make about the operating conditions for this process?

(a)

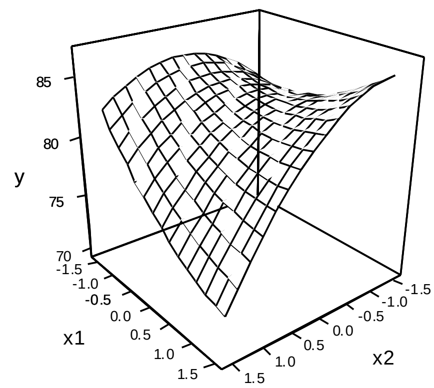


(b)

```
Estimated Regression Coefficients for y
Term             Coef       StDev          T       P
Constant       82.024      0.5622    145.905  0.000
x1             -1.115      0.4397     -2.536  0.044
x2             -2.408      0.4397     -5.475  0.002
x1*x1           0.861      0.7343      1.172  0.286
x2*x2          -1.590      0.7342     -2.165  0.074
x1*x2          -1.801      0.3477     -5.178  0.002
S = 1.390       R-Sq = 92.0%     R-Sq(adj) = 85.3%

Analysis of Variance for y
Source          DF    Seq SS    Adj SS    Adj MS     F      P
Regression       5   132.837   132.837   26.5674  13.74  0.003
  Linear         2    70.393    70.391   35.1957  18.21  0.003
  Square         2    10.602    10.610    5.3048   2.74  0.142
  Interaction    1    51.842    51.842   51.8425  26.82  0.002
Residual Error   6    11.600    11.600    1.9333
  Lack-of-Fit    3    10.052    10.052    3.3507   6.50  0.079
  Pure Error     3     1.548     1.548    0.5158
Total           11   144.437
```

The second order model appears to be significant for the interaction term (p = 0.002). However, the square terms are not significant (p = 0.142).

(c)



There appears to be a saddle point in the experimental region. The yield increases as x_1 is decreased and x_2 is near the zero level.

14-64 An article in *Quality Engineering* ["Mean and Variance Modeling with Qualitative Responses: A Case Study" (1998–1999, Vol. 11, pp. 141–148)] studied how three active ingredients of a particular food affect the overall taste of the product. The measure of the overall taste is the overall mean liking score (MLS). The three ingredients are identified by the variables x_1 , x_2 , and x_3. The data are shown in the following table.

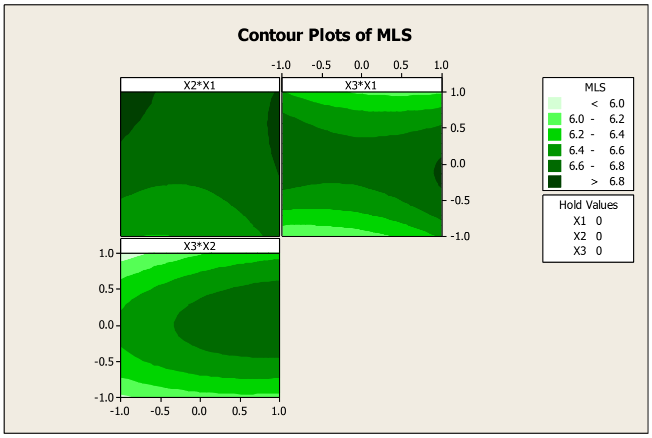
Run	x_1	x_2	x_3	MLS
1	1	1	–1	6.3261
2	1	1	1	6.2444
3	0	0	0	6.5909
4	0	–1	0	6.3409
5	1	–1	1	5.907
6	1	–1	–1	6.488
7	0	0	–1	5.9773
8	0	1	0	6.8605
9	–1	–1	1	6.0455
10	0	0	1	6.3478
11	1	0	0	6.7609
12	–1	–1	–1	5.7727
13	–1	1	–1	6.1805
14	–1	1	1	6.4894
15	–1	0	0	6.8182

(a) Fit a second-order response surface model to the data.
(b) Construct contour plots and response surface plots for MLS. What are your conclusions?
(c) Analyze the residuals from this experiment. Does your analysis indicate any potential problems?
(d) This design has only a single center point. Is this a good design in your opinion?

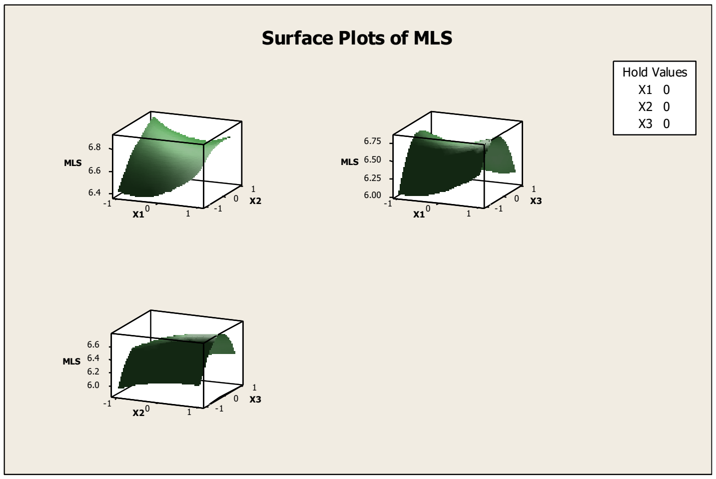
(a)

$$\hat{y} = 6.65821 + 0.04201x_1 + 0.15468x_2 + 0.02895x_3 + 0.11452x_1^2 - 0.07433x_2^2$$
$$- 0.51248x_3^2 - 0.08453x_1x_2 - 0.15555x_1x_3 + 0.06693x_2x_3$$

(b) Contour Plots

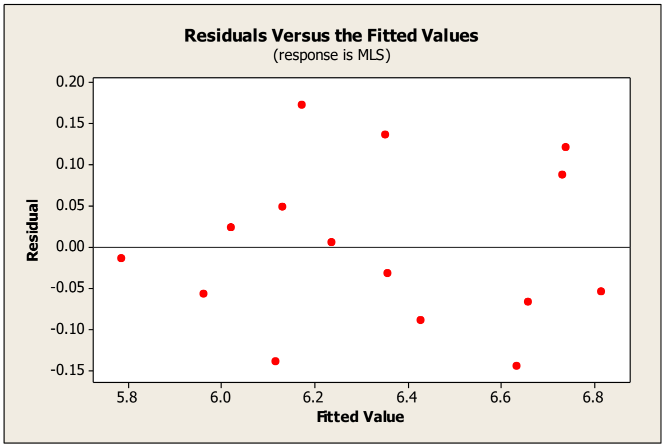


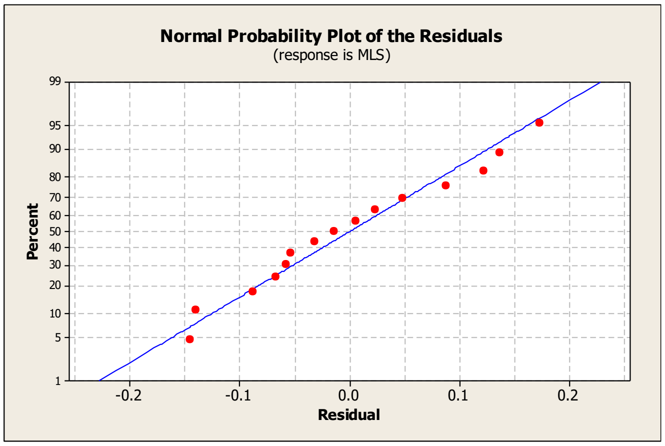
Response surface plots



There is curvature from the second-order effects.

(c) The residual plots appear reasonable.





(d) Adding additional center points would be a good idea to improve the estimates of the coefficients as well as to allow an independent estimate of error to be obtained.

14-65 Consider the first-order model

$$y = 50 + 1.5x_1 - 0.8x_2$$

where $-1 \le x_i \le 1$. Find the direction of steepest ascent.

Move 1.5 units in the direction of x_1 for every –0.8 unit in the direction of x_2. Thus, the path of steepest ascent passes through the point (0, 0) and has a slope -0.8/1.5 = -0.533.

14-66 A manufacturer of cutting tools has developed two empirical equations for tool life (y_1) and tool cost (y_2). Both models are functions of tool hardness (x_1) and manufacturing time (x_2). The equations are

$$y_1 = 10 + 5x_1 + 2x_2$$

$$y_2 = 23 + 3x_1 + 4x_2$$

and both are valid over the range $-1.5 \le x_i \le 1.5$. Suppose that tool life must exceed 12 hours and cost must be below $27.50.

(a) Is there a feasible set of operating conditions?
(b) Where would you run this process?

(a) $10 + 5x_1 + 2x_2 > 12$ $23 + 3x_1 + 4x_2 < 27.50$

$x_2 > -\frac{5}{2}x_1 + 1$ $x_2 < -0.75x_1 + 1.125$

The feasible region is between these two lines, which can be shown graphically on the x_1-x_2 plane.

(b) Operating the process with $x_1 = 1.5$ and $x_2 = -1.5$ results in y_1 and y_2 comfortably within the feasible region.

14-67 An article in *Tappi* (1960, Vol. 43, pp. 38–44) describes an experiment that investigated the ash value of paper pulp (a measure of inorganic impurities). Two variables, temperature *T* in degrees Celsius and time *t* in hours, were studied, and some of the results are shown in the following table. The coded predictor variables shown are

$$x_1 = \frac{(T - 775)}{115}, \qquad x_2 = \frac{(t - 3)}{1.5}$$

and the response *y* is (dry ash value in %) × 103.

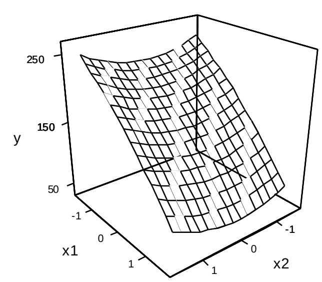
x_1	x_2	y	x_1	x_2	y
–1	–1	211	0	–1.5	168
1	–1	92	0	1.5	179
–1	1	216	0	0	122
1	1	99	0	0	175
–1.5	0	222	0	0	157
1.5	0	48	0	0	146

(a) What type of design has been used in this study? Is the design rotatable?
(b) Fit a quadratic model to the data. Is this model satisfactory?
(c) If it is important to minimize the ash value, where would you run the process?

(a) A central composite design has been used but it is not rotatable.

(b)
```
Term              Coef        StDev          T       P
Constant        150.04        7.821     19.184   0.000
x1              -58.47        5.384    -10.861   0.000
x2                3.35        5.384      0.623   0.556
x1*x1            -6.53        5.693     -1.147   0.295
x2*x2            10.58        5.693      1.859   0.112
x1*x2             0.50        7.848      0.064   0.951
S = 15.70        R-Sq = 95.4%      R-Sq(adj) = 91.6%

Analysis of Variance for y
Source              DF      Seq SS      Adj SS      Adj MS       F       P
Regression           5     30688.7     30688.7      6137.7   24.91   0.001
  Linear             2     29155.4     29155.4     14577.7   59.17   0.000
  Square             2      1532.3      1532.3       766.1    3.11   0.118
  Interaction        1         1.0         1.0         1.0    0.00   0.951
Residual Error       6      1478.2      1478.2       246.4
  Lack-of-Fit        3         4.2         4.2         1.4    0.00   1.000
  Pure Error         3      1474.0      1474.0       491.3
Total               11     32166.9
```

The linear terms appear to be significant (p = 0.001) while both the square terms and interaction terms are insignificant (p = 0.118 and p = 0.951, respectively). Because x_1 is the only significant factor, to minimize *ash* increase the value of x_1.



14-68 In their book *Empirical Model Building and Response Surfaces* (John Wiley, 1987), Box and Draper described an experiment with three factors. The data in the following table are a variation of the original experiment from their book. Suppose that these data were collected in a semiconductor manufacturing process.

(a) The response y_1 is the average of three readings on resistivity for a single wafer. Fit a quadratic model to this response.
(b) The response y_2 is the standard deviation of the three resistivity measurements. Fit a linear model to this response.
(c) Where would you recommend that we set x_1, x_2, and x_3 if the objective is to hold mean resistivity at 500 and minimize the standard deviation?

x_1	x_2	x_3	y_1	y_2
–1	–1	–1	24.00	12.49
0	–1	–1	120.33	8.39
1	–1	–1	213.67	42.83
–1	0	–1	86.00	3.46
0	0	–1	136.63	80.41
1	0	–1	340.67	16.17
–1	1	–1	112.33	27.57
0	1	–1	256.33	4.62
1	1	–1	271.67	23.63
–1	–1	0	81.00	0.00
0	–1	0	101.67	17.67
1	–1	0	357.00	32.91
–1	0	0	171.33	15.01
0	0	0	372.00	0.00
1	0	0	501.67	92.50
–1	1	0	264.00	63.50
0	1	0	427.00	88.61
1	1	0	730.67	21.08
–1	–1	1	220.67	133.82
0	–1	1	239.67	23.46
1	–1	1	422.00	18.52
–1	0	1	199.00	29.44
0	0	1	485.33	44.67
1	0	1	673.67	158.21
–1	1	1	176.67	55.51
0	1	1	501.00	138.94
1	1	1	1010.00	142.45

(a) Response Surface Regression

```
Estimated Regression Coefficients for y

Term          Coef  SE Coef       T      P
Constant    327.62    38.76   8.453  0.000
x3          131.47    17.94   7.328  0.000
x2          109.43    17.94   6.099  0.000
x1          177.00    17.94   9.866  0.000
x3*x3       -29.06    31.08  -0.935  0.363
x2*x2       -22.38    31.08  -0.720  0.481
x1*x1        32.01    31.08   1.030  0.317
x3*x2        43.58    21.97   1.983  0.064
x3*x1        75.47    21.97   3.435  0.003
x2*x1        66.03    21.97   3.005  0.008

S = 76.12    R-Sq = 92.7%    R-Sq(adj) = 88.8%

Analysis of Variance for y

Source            DF    Seq SS    Adj SS   Adj MS      F       P
```

Regression	9	1248237	1248237	138693	23.94	0.000
Linear	3	1090558	1090558	363519	62.74	0.000
Square	3	14219	14219	4740	0.82	0.502
Interaction	3	143461	143461	47820	8.25	0.001
Residual Error	17	98498	98498	5794		
Total	26	1346735				

Reduced model:

Term	Coef	SE Coef	T	P
Constant	314.67	15.46	20.354	0.000
x3	131.47	18.93	6.943	0.000
x2	109.43	18.93	5.779	0.000
x1	177.00	18.93	9.348	0.000
x3*x1	75.47	23.19	3.255	0.004
x2*x1	66.03	23.19	2.847	0.010

S = 80.33   R-Sq = 89.9%   R-Sq(adj) = 87.5%

The quadratic model for y_1 is

$$y_1 = 314.67 + 177.00x_1 + 109.43x_2 + 131.47x_3 + 66.03x_1x_2 + 75.47x_1x_3$$

(b) Response Surface Regression

Estimated Regression Coefficients for y2

Term	Coef	SE Coef	T	P
Constant	48.00	7.808	6.147	0.000
x3	29.19	9.563	3.052	0.006
x2	15.32	9.563	1.602	0.123
x1	11.53	9.563	1.205	0.240

S = 40.57   R-Sq = 36.7%   R-Sq(adj) = 28.4%

Analysis of Variance for y2

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	3	21957.3	21957.3	7319.09	4.45	0.013
Linear	3	21957.3	21957.3	7319.09	4.45	0.013
Residual Error	23	37863.6	37863.6	1646.24		
Total	26	59820.9				

The linear model for y_2 is given by

$$y_2 = 48.00 + 29.19x_3$$

(c) The equations for y_1 and y_2 are used to determine values for the x's. Given values for x_1 and x_2, a value for x_3 can be solved to set y_1 to a target. Each x_i should range from -1 to 1 to stay within the experimental region for the models. The standard deviation is minimized with the smallest feasible value for x_3. When x_3 = -1 at least one of x_1 and x_2 must exceed 1 in order to set y_1 = 500. Therefore, x_3 is greater than -1. To keep a solution within the feasible region, we set x_1 = 1 and x_2 = 1. With these values the value for x_3 that sets y_1 = 500 is x_3 = -0.808 and this minimizes the standard deviation.

14-69 Consider the first-order model

$$y = 12 + 1.2x_1 - 2.1x_2 + 1.6x_3 - 0.6x_4$$

where $-1 \le x_i \le 1.$

(a) Find the direction of steepest ascent.
(b) Assume that the current design is centered at the point (0, 0, 0, 0). Determine the point that is three units from the current center point in the direction of steepest ascent.

(a) $y = 12 + 1.2x_1 - 2.1x_2 + 1.6x_3 - 0.6x_4$

The direction of steepest ascent is in the direction of the vector (1.2, -2.1, 1.6, -0.6)

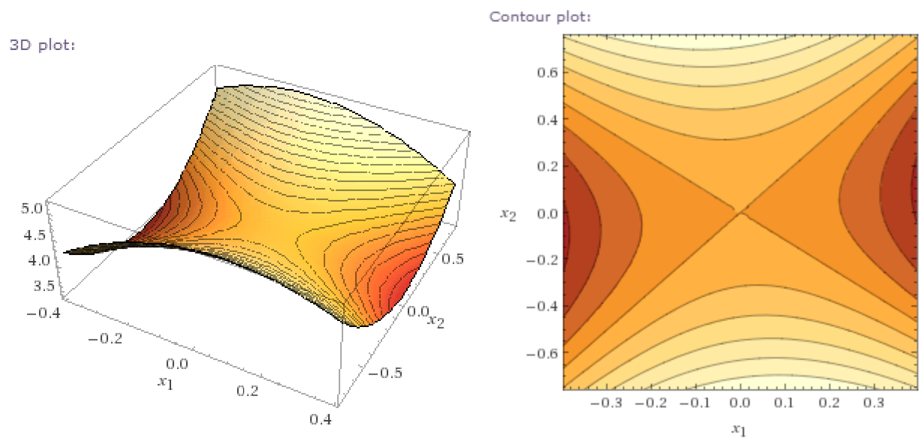
(b) The point along the path of steepest descent that is 3 units away from (0,0,0,0) is given by:

$$\frac{3\cdot(1.2,\quad -2.1,\quad 1.6,\quad -0.6)}{\sqrt{1.2^2+(-2.1)^2+1.6^2+(-0.6)^2}}=\frac{3\cdot(1.2,\quad -2.1,\quad 1.6,\quad -0.6)}{2.9614}=\left(1.22,\quad -2.13,\quad 1.62,\quad -0.61\right)$$

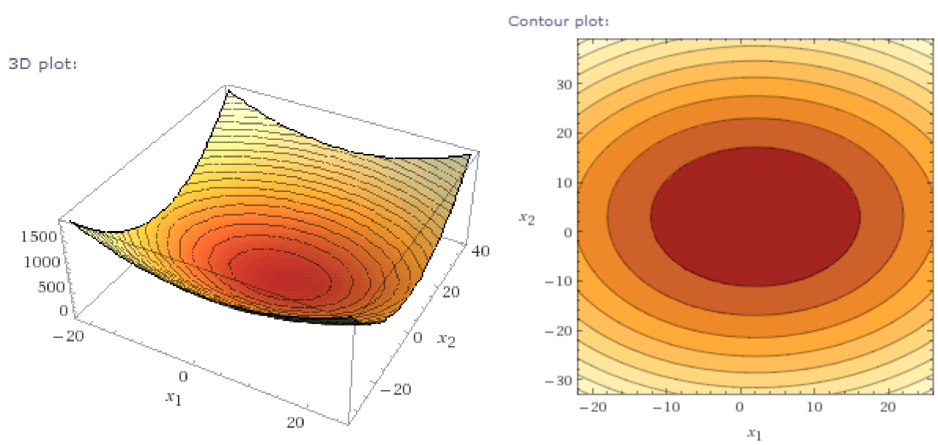
14-70 Suppose that a response y_1 is a function of two inputs x_1 and x_2 with $y_1 = 2x_2^2 - 4x_1^2 - x_1x_2 + 4$.

(a) Draw the contours of this response function.
(b) Consider another response $y_2 = (x_1 - 2)^2 + (x_2 - 3)^2$.
(c) Add the contours for $y2$ and discuss how feasible it is to minimize both y_1 and y_2 with values for x_1 and x_2.

(a) $y_1 = 2x_2^2 - 4x_1^2 - x_1x_2 + 4$



(b) $y_2 = (x_1 - 2)^2 + (x_2 - 3)^2$

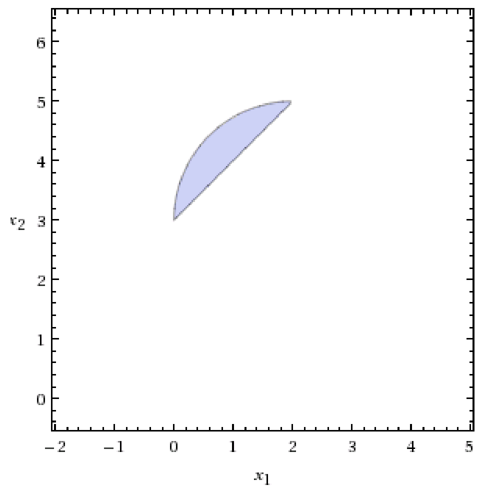


(c) It is impossible to minimize y_1, and y_2 is minimized at x_1=2 and x_2 =3

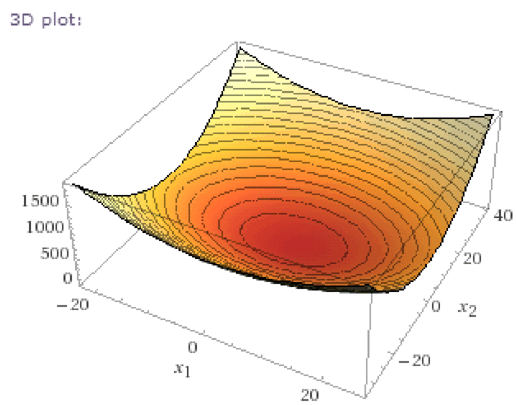
14-71 Two responses $y1$ and $y2$ are related to two inputs x_1 and x_2 by the models $y_1 = 5 + (x_1 - 2)^2 + (x_2 - 3)^2$ and $y_2 = x_2 - x_1 + 3$. Suppose that the objectives are $y_1 \le 9$ and $y_2 \ge 6$.

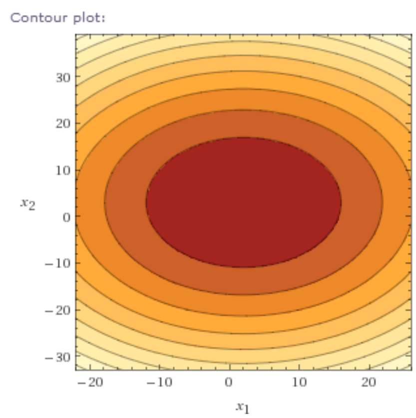
(a) Is there a feasible set of operating conditions for x_1 and x_2? If so, plot the feasible region in the space of x_1 and x_2.
(b) Determine the point(s) (x_1, x_2) that yields $y_2 \ge 6$ and minimizes y_1.

(a)The region $y_1 < 9$ is a circle in (x_1, x_2) space centered as the point (2,3) with radius 2. The region $y2 > 6$ is the half plane $x_2 > x_1 + 3$. The following graph shows the feasible region.



(b) Because $y_1 = 5 + (x_1 - 2)^2 + (x_2 - 3)^2 \leq 9$





The minimum is found in ($x_1 = 1, x_2 = 4$) with $y_1 = 7$

14-72 An article in the *Journal of Materials Processing Technology* (1997, Vol. 67, pp. 55–61) used response surface methodology to generate surface roughness prediction models for turning EN 24T steel (290 BHN). The data are shown in the following table.

Trial	Speed (m min^{-1})	Feed (mm rev^{-1})	Depth of cut (mm)	Coding x_1	x_2	x_3	Surface roughness, (μm)
1	36	0.15	0.50	–1	–1	–1	1.8
2	117	0.15	0.50	1	–1	–1	1.233
3	36	0.40	0.50	–1	1	–1	5.3
4	117	0.40	0.50	1	1	–1	5.067
5	36	0.15	1.125	–1	–1	1	2.133
6	117	0.15	1.125	1	–1	1	1.45
7	36	0.40	1.125	–1	1	1	6.233
8	117	0.40	1.125	1	1	1	5.167
9	65	0.25	0.75	0	0	0	2.433
10	65	0.25	0.75	0	0	0	2.3
11	65	0.25	0.75	0	0	0	2.367
12	65	0.25	0.75	0	0	0	2.467
13	28	0.25	0.75	$-\sqrt{2}$	0	0	3.633
14	150	0.25	0.75	$\sqrt{2}$	0	0	2.767
15	65	0.12	0.75	0	$-\sqrt{2}$	0	1.153
16	65	0.50	0.75	0	$\sqrt{2}$	0	6.333
17	65	0.25	0.42	0	0	$-\sqrt{2}$	2.533
18	65	0.25	1.33	0	0	$\sqrt{2}$	3.20
19	28	0.25	0.75	$-\sqrt{2}$	0	0	3.233
20	150	0.25	0.75	$\sqrt{2}$	0	0	2.967
21	65	0.12	0.75	0	$-\sqrt{2}$	0	1.21
22	65	0.50	0.75	0	$\sqrt{2}$	0	6.733
23	65	0.25	0.42	0	0	$-\sqrt{2}$	2.833
24	65	0.25	1.33	0	0	$\sqrt{2}$	3.267

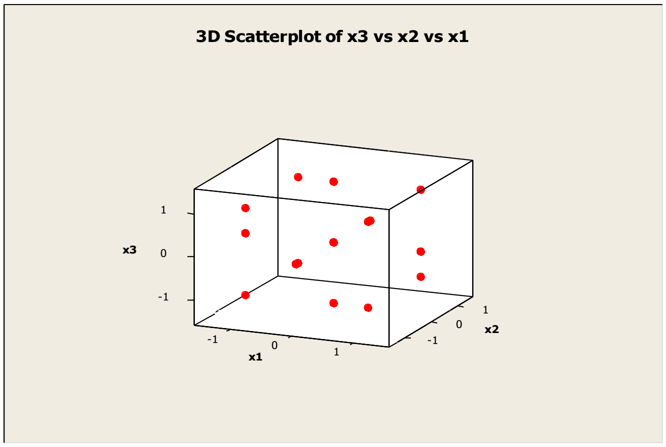
The factors and levels for the experiment are shown in Table E14-3.

TABLE • E14-3 Steel Factors

Levels	Lowest	Low	Center	High	Highest
Coding	$-\sqrt{2}$	–1	0	1	$\sqrt{2}$
Speed, V (m min^{-1})	28	36	65	117	150
Feed, f (mm rev $^{-1}$)	0.12	0.15	0.25	0.40	0.50
Depth of cut, d (mm)	0.42	0.50	0.75	1.125	1.33

(a) Plot the points at which the experimental runs were made.
(b) Fit both first-and second-order models to the data. Comment on the adequacies of these models.
(c) Plot the roughness response surface for the second-order model and comment.

(a) A plot of the *coded* data follows.



(b) Computer results are shown below for the first-order and second-order models for the *coded* data. Note that the coded data are computed after a natural logarithm transform is applied to the original data. That is, the center point for the variable *speed* is not (117+36)/2 = 76.5, but instead [ln(117) + ln(36)]/2 = 4.1728 and exp(4.1728) = 65.

Response Surface Regression: y versus x1, x2, x3

The analysis was done using coded units.

Estimated Regression Coefficients for y

Term	Coef	SE Coef	T	P
Constant	3.2422	0.1120	28.955	0.000
x1	-0.2594	0.1371	-1.891	0.073
x2	1.8931	0.1371	13.804	0.000
x3	0.1963	0.1371	1.431	0.168

S = 0.5485   R-Sq = 90.7%   R-Sq(adj) = 89.4%

Analysis of Variance for y

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	3	59.0253	59.0253	19.6751	65.39	0.000
Linear	3	59.0253	59.0253	19.6751	65.39	0.000
Residual Error	20	6.0180	6.0180	0.3009		
Lack-of-Fit	11	5.7727	5.7727	0.5248	19.26	0.000
Pure Error	9	0.2453	0.2453	0.0273		
Total	23	65.0432				

Note that the lack-of-fit test is significant for the first-order model (P-value near zero) and this indicates that a second-order model should be considered.

Response Surface Regression: Surface Roughness versus x1, x2, x3

The analysis was done using coded units.

Estimated Regression Coefficients for Surface Roughness

Term	Coef	SE Coef	T	P
Constant	2.47142	0.08780	28.147	0.000
x1	-0.25937	0.04809	-5.393	0.000
x2	1.89296	0.04809	39.361	0.000
x3	0.19625	0.04809	4.081	0.001
x1*x1	0.29946	0.05553	5.393	0.000
x2*x2	0.65308	0.05553	11.760	0.000
x3*x3	0.20358	0.05553	3.666	0.003
x1*x2	-0.00612	0.06801	-0.090	0.930

x1*x3	-0.11863	0.06801	-1.744	0.103
x2*x3	0.06038	0.06801	0.888	0.390

S = 0.1924 R-Sq = 99.2% R-Sq(adj) = 98.7%

Analysis of Variance for Surface Roughness

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	9	64.5252	64.5252	7.1695	193.74	0.000
Linear	3	59.0252	59.0252	19.6751	531.67	0.000
Square	3	5.3579	5.3579	1.7860	48.26	0.000
Interaction	3	0.1420	0.1420	0.0473	1.28	0.320
Residual Error	14	0.5181	0.5181	0.0370		
Lack-of-Fit	5	0.2728	0.2728	0.0546	2.00	0.172
Pure Error	9	0.2453	0.2453	0.0273		
Total	23	65.0432				

The linear and pure quadratic terms appear to be significant (P-value = 0 and P-value = 0) while the interaction terms are insignificant (P-value = 0.32). The lack-of-fit test is not significant and this indicates a better fit for the second-order model.

Reduced model

Response Surface Regression: Roughness versus x1, x2, x3

The analysis was done using coded units.

Estimated Regression Coefficients for Roughness

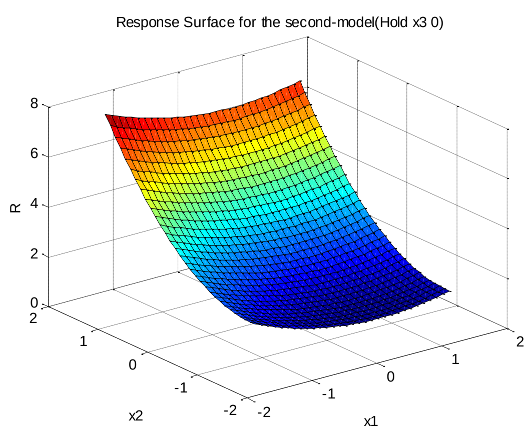
Term	Coef	SE Coef	T	P
Constant	2.4714	0.08994	27.478	0.000
x1	-0.2594	0.04926	-5.265	0.000
x2	1.8930	0.04926	38.425	0.000
x3	0.1963	0.04926	3.984	0.001
x1*x1	0.2995	0.05688	5.264	0.000
x2*x2	0.6531	0.05688	11.481	0.000
x3*x3	0.2036	0.05688	3.579	0.002

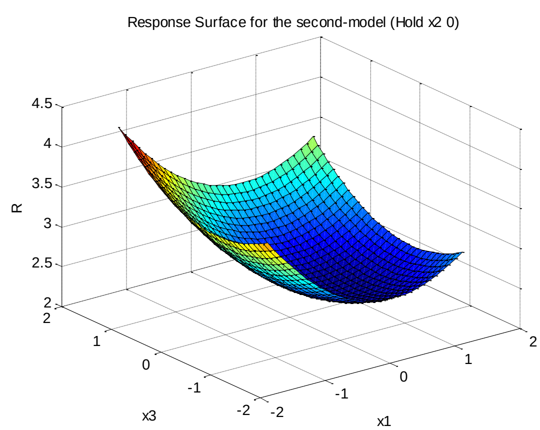
S = 0.1971 R-Sq = 99.0% R-Sq(adj) = 98.6%

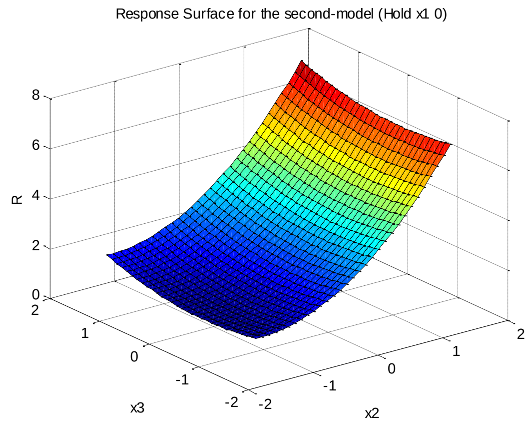
The quadratic model of the coded variable is

$$y_1 = 2.4714 - 0.2594x_1 + 1.8930x_2 + 0.1963x_3 + 0.2995x_1^2 + 0.6531x_2^2 + 0.2036x_3^2$$

(c) There is curvature in the fitted surface from the second-order effects.







14-73 An article in *Analytical Biochemistry* ["Application of Central Composite Design for DNA Hybridization Onto Magnetic Microparticles," (2009, Vol. 391(1), 2009, pp. 17–23)] considered the effects of probe and target concentration and particle number in immobilization and hybridization on a micro particle-based DNA hybridization assay. Mean fluorescence is the response. Particle concentration was transformed to surface area measurements. Other concentrations were measured in micromoles per liter (μM). Data are in Table E14-2.

TABLE • E14-2 Fluorescence Experiment

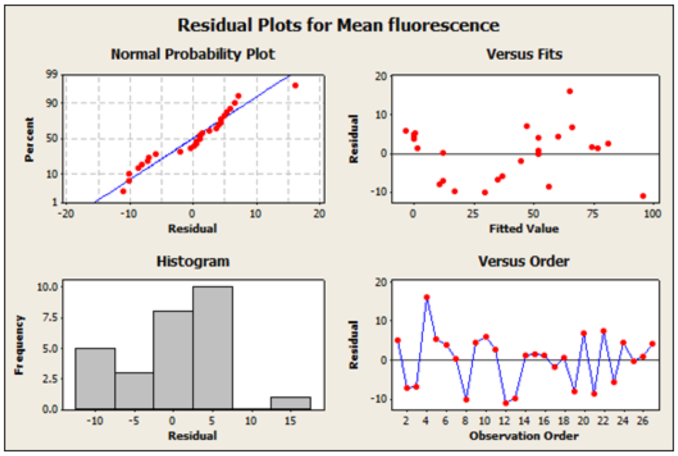
Run	Immobilization area (cm^2)	Probe area (μM)	Hybridization (cm^2)	Target (μM)	Mean Fluorescence
1	0.35	0.025	0.35	0.025	4.7
2	7	0.025	0.35	0.025	4.7
3	0.35	2.5	0.35	0.025	28.0
4	7	2.5	0.35	0.025	81.2
5	0.35	0.025	3.5	0.025	5.7
6	7	0.025	3.5	0.025	3.8
7	0.35	2.5	3.5	0.025	12.2
8	7	2.5	3.5	0.025	19.5
9	0.35	0.025	0.35	5	4.4
10	7	0.025	0.35	5	2.6
11	0.35	2.5	0.35	5	83.7
12	7	2.5	0.35	5	84.7
13	0.35	0.025	3.5	5	6.8
14	7	0.025	3.5	5	2.4
15	0.35	2.5	3.5	5	76
16	7	2.5	3.5	5	77.9
17	0.35	5	2	2.5	42.6
18	7	5	2	2.5	52.3
19	3.5	0.025	2	2.5	2.6
20	3.5	2.5	2	2.5	72.8
21	3.5	5	0.35	2.5	47.7
22	3.5	5	3.5	2.5	54.4
23	3.5	5	2	0.025	30.8
24	3.5	5	2	5	64.8
25	3.5	5	2	2.5	51.6
26	3.5	5	2	2.5	52.6
27	3.5	5	2	2.5	56.1

(a) What type of design is used?
(b) Fit a second-order response surface model to the data.
(c) Does a residual analysis indicate any problems?

(a) A central composite design is used with four factors and $\alpha = 1$

(b)
```
The analysis was done using coded units.

Estimated Regression Coefficients for Mean fluorescence

Term                                 Coef  SE Coef       T      P
Constant                          51.9630    3.589  14.479  0.000
Immobilization area                3.6111    2.295   1.573  0.142
Probe area                        27.6833    2.295  12.060  0.000
Hybridization                     -4.6111    2.295  -2.009  0.068
Target                            11.8167    2.295   5.148  0.000
Immobilization area*              -3.7778    6.073  -0.622  0.546
  Immobilization area
Probe area*Probe area            -13.5278    6.073  -2.227  0.046
Hybridization*Hybridization       -0.1778    6.073  -0.029  0.977
Target*Target                     -3.4278    6.073  -0.564  0.583
Immobilization area*Probe area     4.4688    2.435   1.835  0.091
Immobilization area*Hybridization -3.0937    2.435  -1.271  0.228
Immobilization area*Target        -3.8687    2.435  -1.589  0.138
Probe area*Hybridization          -5.8938    2.435  -2.421  0.032
Probe area*Target                 11.5062    2.435   4.726  0.000
Hybridization*Target               4.0688    2.435   1.671  0.121


S = 9.73866    PRESS = 8316.94
R-Sq = 95.23%  R-Sq(pred) = 65.17%  R-Sq(adj) = 89.67%


Analysis of Variance for Mean fluorescence

Source                                   DF   Seq SS   Adj SS   Adj MS
Regression                               14  22743.7  22743.7   1624.6
  Linear                                  4  16925.5  16925.5   4231.4
    Immobilization area                   1    234.7    234.7    234.7
    Probe area                            1  13794.6  13794.6  13794.6
    Hybridization                         1    382.7    382.7    382.7
    Target                                1   2513.4   2513.4   2513.4
  Square                                  4   2167.2   2167.2    541.8
    Immobilization area*Immobilization area  1   1386.2     36.7     36.7
    Probe area*Probe area                 1    747.0    470.6    470.6
    Hybridization*Hybridization           1      3.7      0.1      0.1
    Target*Target                         1     30.2     30.2     30.2
  Interaction                             6   3651.1   3651.1    608.5
    Immobilization area*Probe area        1    319.5    319.5    319.5
    Immobilization area*Hybridization     1    153.1    153.1    153.1
    Immobilization area*Target            1    239.5    239.5    239.5
    Probe area*Hybridization              1    555.8    555.8    555.8
    Probe area*Target                     1   2118.3   2118.3   2118.3
    Hybridization*Target                  1    264.9    264.9    264.9
Residual Error                           12   1138.1   1138.1     94.8
  Lack-of-Fit                            10   1126.9   1126.9    112.7
  Pure Error                              2     11.2     11.2      5.6
Total                                    26  23881.8
```

(c) The residual plots do not indicate any problems.



14-74 An article in *Applied Biochemistry and Biotechnology*("A Statistical Approach for Optimization of Polyhydroxybutyrate Production by Bacillus sphaericus NCIM 5149 under Submerged Fermentation Using Central Composite Design" (2010, Vol. 162(4), pp. 996–1007)] described an experiment to optimize the production of polyhydroxybutyrate (PHB). Inoculum age, pH, and substrate were selected as factors, and a central composite design was conducted. Data follow.

Run	Inoculum age (h)	pH	Substrate (g/L)	PHB (g/L)
1	12	4	1	0.84
2	24	8	1	0.55
3	18	6	2.5	1.96
4	28	6	2.5	1.2
5	12	4	4	0.783
6	18	6	2.5	1.66
7	18	6	2.5	2.22
8	18	6	5	0.8
9	12	8	4	0.48
10	18	6	2.5	1.97
11	18	6	2.5	2.2
12	18	6	2.5	2.25
13	18	2	2.5	0.2
14	18	6	0	0.22
15	12	8	1	0.37
16	24	8	4	0.66
17	24	4	1	0.28
18	24	4	4	0.88
19	18	9	2.5	0.3
20	7	6	2.5	0.42

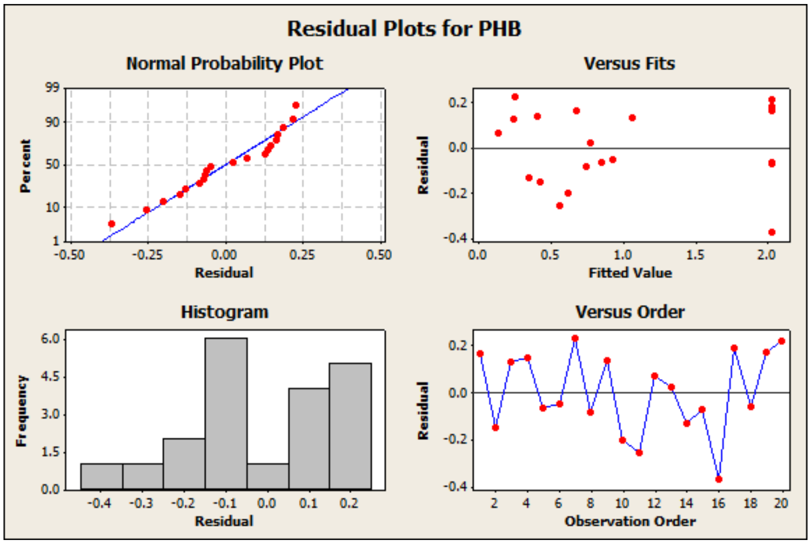
(a) Plot the points at which the experimental runs were made [*Hint*: Code each variable first.] What type of design is used?
(b) Fit a second-order response surface model to the data.
(c) Does a residual analysis indicate any problems?
(d) Construct a contour plot and response surface for PHB amount in terms of two factors.
(e) Can you recommend values for inoculum age, pH and substrate to maximize production?

(a) Central composite design with three factors and alpha = 1.68

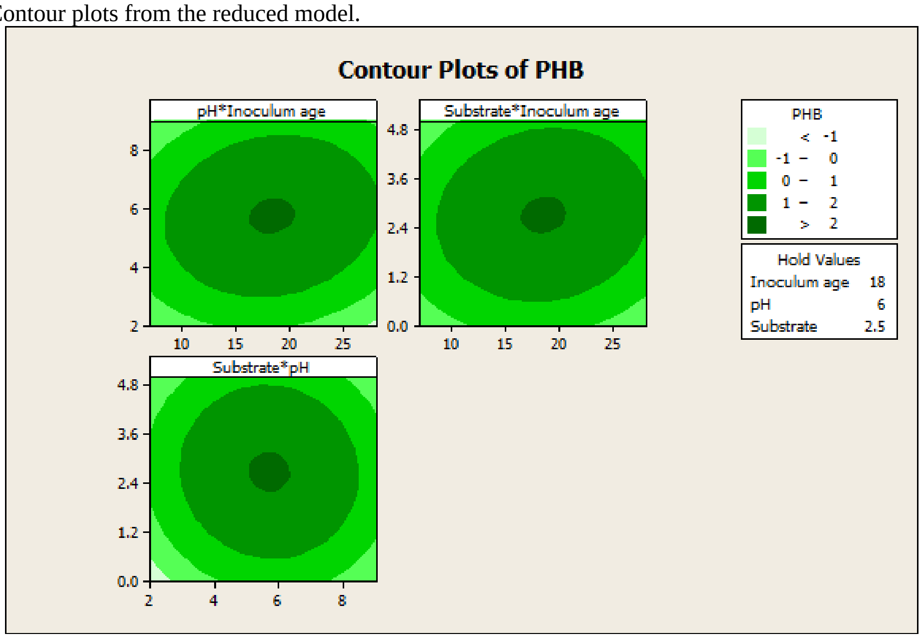
(b)

```
The analysis was done using coded units.

Estimated Regression Coefficients for PHB

Term                              Coef  SE Coef       T      P
Constant                       2.03328  0.09577  21.232  0.000
Inoculum age                   0.06188  0.06316   0.980  0.350
pH                            -0.15527  0.06491  -2.392  0.038
Substrate                      0.12760  0.06423   1.987  0.075
Inoculum age*Inoculum age     -0.38565  0.05834  -6.610  0.000
pH*pH                         -0.55270  0.05704  -9.689  0.000
Substrate*Substrate           -0.52830  0.06340  -8.333  0.000
Inoculum age*pH                0.10288  0.08361   1.230  0.247
Inoculum age*Substrate         0.08213  0.08361   0.982  0.349
pH*Substrate                  -0.04037  0.08361  -0.483  0.640


S = 0.236489   PRESS = 2.69564
R-Sq = 94.69%  R-Sq(pred) = 74.40%  R-Sq(adj) = 89.91%

Analysis of Variance for PHB

Source                         DF   Seq SS   Adj SS   Adj MS      F      P
Regression                      9   9.9725  9.97247  1.10805  19.81  0.000
  Linear                        3   0.3414  0.59413  0.19804   3.54  0.056
    Inoculum age                1   0.1187  0.05368  0.05368   0.96  0.350
    pH                          1   0.0019  0.31999  0.31999   5.72  0.038
    Substrate                   1   0.2207  0.22070  0.22070   3.95  0.075
  Square                        3   9.4794  9.47940  3.15980  56.50  0.000
    Inoculum age*Inoculum age   1   1.2484  2.44389  2.44389  43.70  0.000
    pH*pH                       1   4.3477  5.25009  5.25009  93.87  0.000
    Substrate*Substrate         1   3.8832  3.88321  3.88321  69.43  0.000
  Interaction                   3   0.1517  0.15166  0.05055   0.90  0.473
    Inoculum age*pH             1   0.0847  0.08467  0.08467   1.51  0.247
    Inoculum age*Substrate      1   0.0540  0.05396  0.05396   0.96  0.349
    pH*Substrate                1   0.0130  0.01304  0.01304   0.23  0.640
Residual Error                 10   0.5593  0.55927  0.05593
  Lack-of-Fit                   5   0.3015  0.30154  0.06031   1.17  0.434
  Pure Error                    5   0.2577  0.25773  0.05155
Total                          19  10.5317
Estimated Regression Coefficients for PHB using data in uncoded units

Term                              Coef
Constant                      -6.67722
Inoculum age                  0.321711
pH                             1.45980
Substrate                      1.17557
Inoculum age*Inoculum age   -0.0107124
pH*pH                        -0.138175
Substrate*Substrate          -0.234800
Inoculum age*pH             0.00857292
Inoculum age*Substrate      0.00912500
pH*Substrate                -0.0134583
```

(c) Residual plots for the reduced model.



(d) Contour plots from the reduced model.



(e) Based on the contour plots, values for pH near 6, substrate near 2.5 and Inoculum age near 20 are recommended.

Supplemental Exercises

14-75 An article in *Process Engineering* (1992, No. 71, pp. 46–47) presented a two-factor factorial experiment to investigate the effect of pH and catalyst concentration on product viscosity (cSt). The data are as follows:

	Catalyst Concentration	
pH	2.5	2.7
5.6	192, 199, 189, 198	178, 186, 179, 188
5.9	185, 193, 185, 192	197, 196, 204, 204

(a) Test for main effects and interactions using $\alpha = 0.05$. What are your conclusions?
(b) Graph the interaction and discuss the information provided by this plot.

(c) Analyze the residuals from this experiment.

(a) Estimated Effects and Coefficients for var_1 (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		191.563	1.158	165.49	0.000
factor_A (PH)	5.875	2.937	1.158	2.54	0.026
factor_B (CC)	-0.125	-0.062	1.158	-0.05	0.958
factor_A*factor_B	11.625	5.812	1.158	5.02	0.000

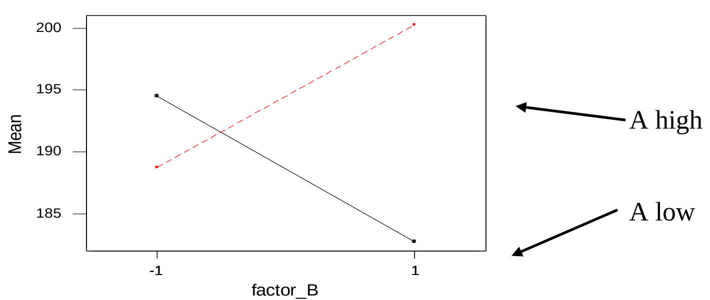
Analysis of Variance for var_1 (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	2	138.125	138.125	69.06	3.22	0.076
2-Way Interactions	1	540.562	540.562	540.56	25.22	0.000
Residual Error	12	257.250	257.250	21.44		
Pure Error	12	257.250	257.250	21.44		
Total	15	935.938				

The main effect of *pH* and the interaction of *pH*Catalyst Concentration* (CC) are significant at the 0.05 level of significance. The model used is viscosity = $191.563 + 2.937x_1 - 0.062x_2 + 5.812x_{12}$

(b) The interaction plot shows that there is a strong interaction. When Factor A is at its low level, the mean response is large at the low level of B and is small at the high level of B. However, when A is at its high level, the results reverse.

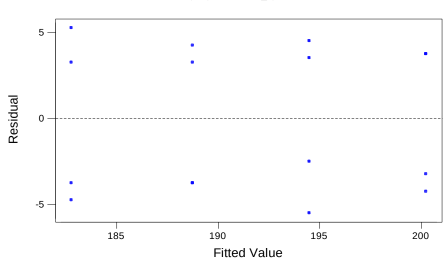
Interaction Plot (data means) for var_1

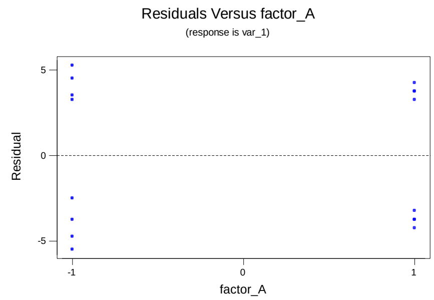


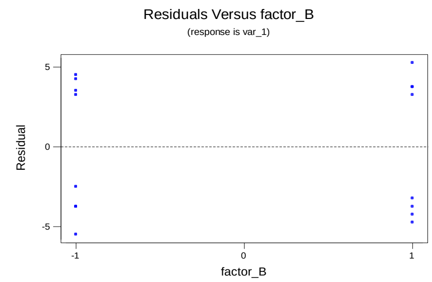
(c) The plots of the residuals show that the equality of variance assumption is reasonable. However, there is a large gap in the middle of the normal probability plot. Sometimes, this can indicate that there is another variable that has an effect on the response, but which is not included in the experiment. For example, in this experiment, note that the replicates in each cell have two pairs of values that are very similar, but there is a rather large difference in the mean values of the two pairs. (Cell 1 has 189 and 192 as one pair and 198 and 199 as the other.)

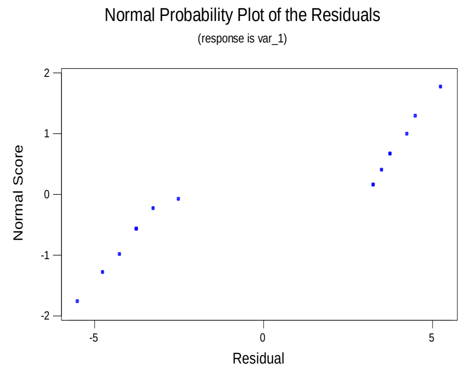
Residuals Versus the Fitted Values

(response is var_1)









14-76 Heat-treating metal parts is a widely used manufacturing process. An article in the *Journal of Metals* (1989, Vol. 41) described an experiment to investigate flatness distortion from heat-treating for three types of gears and two heat-treating times. The data follow:

	Time (minutes)	
Gear Type	**90**	**120**
20-tooth	0.0265	0.0560
	0.0340	0.0650
24-tooth	0.0430	0.0720
	0.0510	0.0880
28-tooth	0.0405	0.0620
	0.0575	0.0825

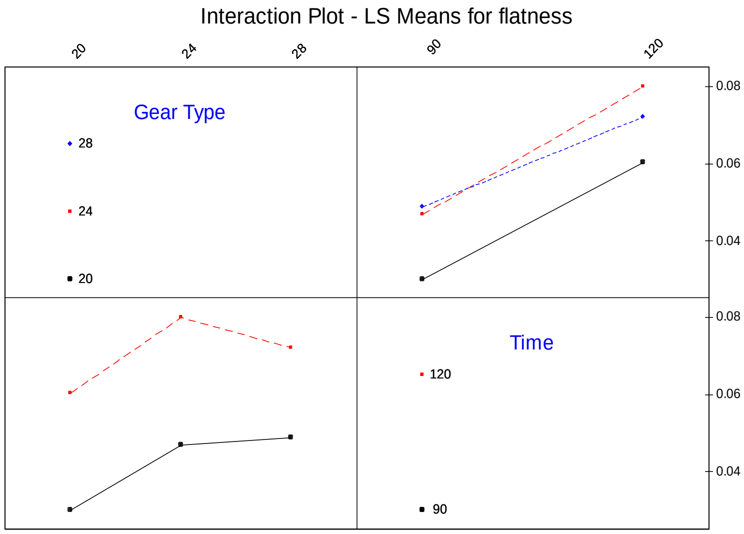
(a) Is there any evidence that flatness distortion is different for the different gear types? Is there any indication that heat-treating time affects the flatness distortion? Do these factors interact? Use $\alpha = 0.05$.
(b) Construct graphs of the factor effects that aid in drawing conclusions from this experiment.
(c) Analyze the residuals from this experiment. Comment on the validity of the underlying assumptions.

(a)
```
Factor      Type Levels Values
Gear Typ   fixed      3 20 24 28
Time       fixed      2  90 120

Analysis of Variance for flatness, using Adjusted SS for Tests

Source          DF     Seq SS     Adj SS     Adj MS       F      P
Gear Typ         2  0.0007591  0.0007591  0.0003796    3.90  0.082
Time             1  0.0024941  0.0024941  0.0024941   25.66  0.002
Gear Typ*Time    2  0.0000505  0.0000505  0.0000253    0.26  0.779
Error            6  0.0005833  0.0005833  0.0000972
Total           11  0.0038870
```

Term		Coef	SE Coef	T	P
Constant		0.056500	0.002846	19.85	0.000
Gear Typ					
20		-0.011125	0.004025	-2.76	0.033
24		0.007000	0.004025	1.74	0.133
Time					
90		-0.014417	0.002846	-5.07	0.002
Gear Typ*Time					
20	90	-0.000708	0.004025	-0.18	0.866
24	90	-0.002083	0.004025	-0.52	0.623

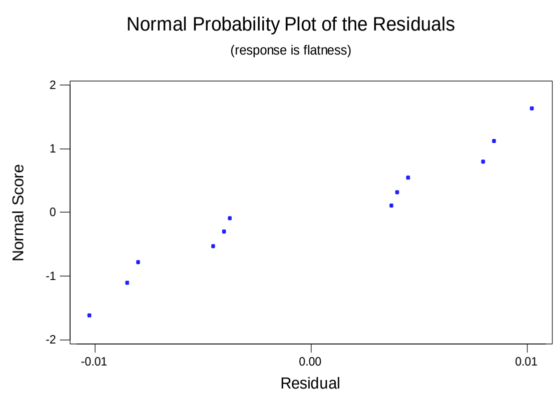
There is weak evidence that flatness distortion is different for the different gear types (p = 0.082). Gear type is significant at α = 0.1, but not at α = 0.05. Also, the gear type 20 coefficient has a p-value = 0.033. Heat-treating time affects the flatness distortion (p = 0.002). There is no evidence that factors interact (p = 0.779).

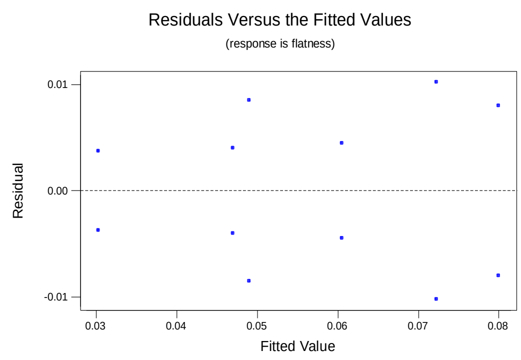
(b)

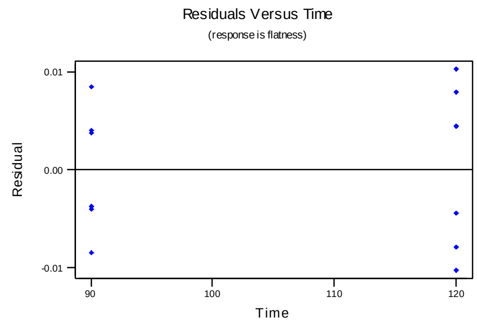


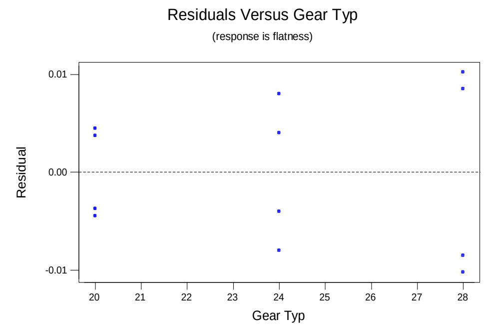
The interaction plot for the effects indicates that there is no interaction between gear type and time. The interaction plot indicates there may be some significant difference between the low and high levels of time. There is also a difference between one of the gear types and the other two.

(c) The model used is $\hat{y} = 0.0565 = 0.0111x_1 - 0.01144x_2$









The residual plots are adequate. There does not appear to be any serious departure from normality or violation of the assumption of constant variance.

14-77 An article in the *Textile Research Institute Journal* (1984, Vol. 54, pp. 171–179) reported the results of an experiment that studied the effects of treating fabric with selected inorganic salts on the flammability of the material. Two application levels of each salt were used, and a vertical burn test was used on each sample. (This finds the temperature at which each sample ignites.) The burn test data follow.

	Salt					
Level	**Untreated**	**$MgCl_2$**	**NaCl**	**$CaCO_3$**	**$CaCl_2$**	**Na_2CO_3**
1	812	752	739	733	725	751
	827	728	731	728	727	761
	876	764	726	720	719	755
2	945	794	741	786	756	910
	881	760	744	771	781	854
	919	757	727	779	814	848

(a) Test for differences between salts, application levels, and interactions. Use α = 0.01.
(b) Draw a graph of the interaction between salt and application level. What conclusions can you draw from this graph?
(c) Analyze the residuals from this experiment.

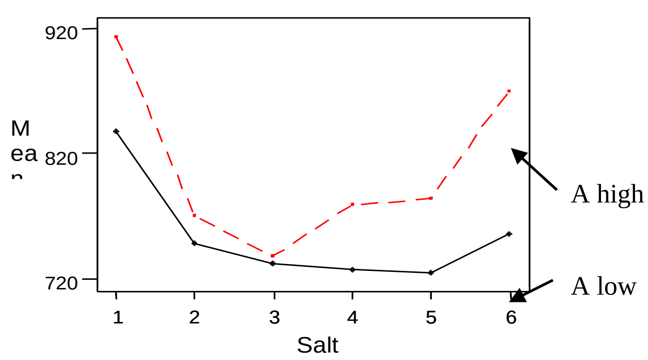
(a)
```
Factor      Type   Levels  Values
 Level      fixed      2      1      2
 Salt       fixed      6      1      2      3      4      5      6

Analysis of Variance for temperat
Source          DF         SS         MS         F         P
level            1      27390      27390     63.24     0.000
salt             5      86087      17217     39.75     0.000
```

Interaction	5	11459	2292	5.29	0.002
Error	24	10395	433		
Total	35	135332			

There is a significant difference between the application levels, the salts, and there is a significant interaction between the two factors.

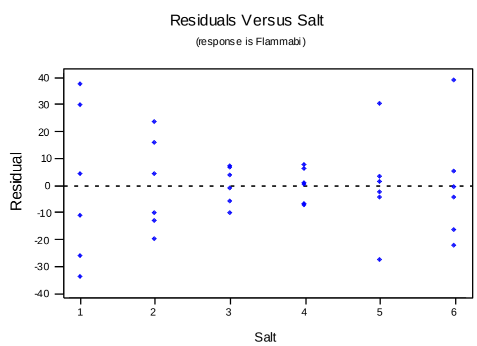
(b)

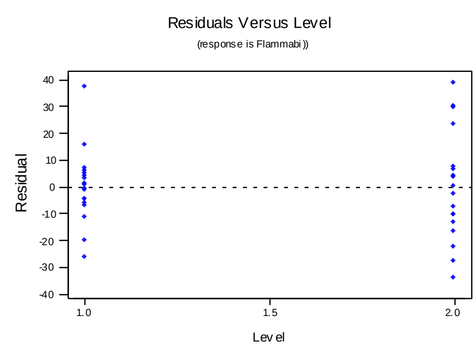
Interaction Plot - Means for Temperature

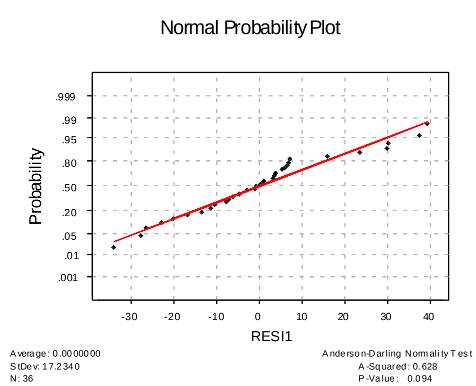


From the interaction plot, we see that the untreated salt 1 has a higher flammability average than any of the other five levels. The remaining five levels ($MgCl_2$, NaCl, $CaCO_3$, $CaCl_2$, Na_2CO_3) also seem to differ. Overall, application level 1 increases the flammability average. Also, the difference between application levels varies with the salt type and this indicates a significant interaction.

(c)







The residual plots do not indicate major problems with the assumptions. There is some concern with the constant variability assumption in the plot of residuals versus salts.

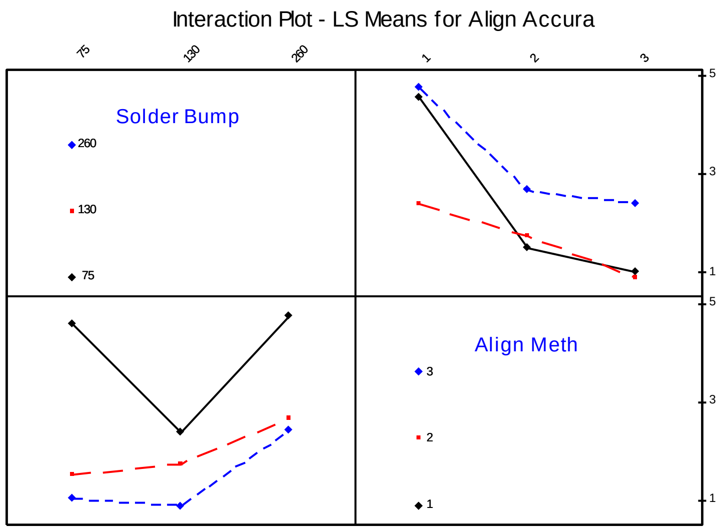
14-78 An article in the *IEEE Transactions on Components, Hybrids, and Manufacturing Technology* (1992, Vol. 15) described an experiment for aligning optical chips onto circuit boards. The method involves placing solder bumps onto the bottom of the chip. The experiment used three solder bump sizes and three alignment methods. The response variable is alignment accuracy (in micrometers). The data are as follows:

Solder Bump Size (diameter in mm)	Alignment Method 1	2	3
	4.60	1.55	1.05
75	4.53	1.45	1.00
	2.33	1.72	0.82
130	2.44	1.76	0.95
	4.95	2.73	2.36
260	4.55	2.60	2.46

(a) Is there any indication that either solder bump size or alignment method affects the alignment accuracy? Is there any evidence of interaction between these factors? Use $\alpha = 0.05$.
(b) What recommendations would you make about this process?
(c) Analyze the residuals from this experiment. Comment on model adequacy.

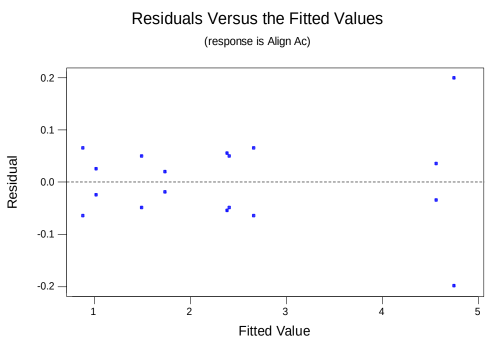
(a)
```
Factor     Type Levels Values
Solder B  fixed      3  75 130 260
Align Me  fixed      3 1 2 3

Analysis of Variance for Align Ac, using Adjusted SS for Tests

Source               DF     Seq SS     Adj SS     Adj MS       F       P
Solder B              2     7.7757     7.7757     3.8879  297.92   0.000
Align Me              2    20.1241    20.1241    10.0621  771.04   0.000
Solder B*Align Me     4     3.5001     3.5001     0.8750   67.05   0.000
Error                 9     0.1174     0.1174     0.0130
Total                17    31.5174

Term                    Coef   SE Coef        T       P
Constant             2.43611   0.02693    90.47   0.000
Solder B
 75                 -0.07278   0.03808    -1.91   0.088
130                 -0.76611   0.03808   -20.12   0.000
Align Me
1                    1.46389   0.03808    38.44   0.000
2                   -0.46778   0.03808   -12.28   0.000
Solder B*Align Me
 75      1           0.73778   0.05385    13.70   0.000
 75      2          -0.39556   0.05385    -7.35   0.000
130      1          -0.74889   0.05385   -13.91   0.000
130      2           0.53778   0.05385     9.99   0.000
```

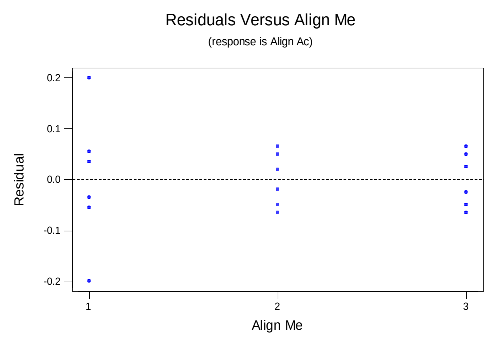
The analysis indicates that both solder size and alignment method significantly affect alignment accuracy. The interaction between solder size and alignment method is also significant in affecting alignment accuracy.
(b) The lines for factor A intersect at the lower level of alignment.

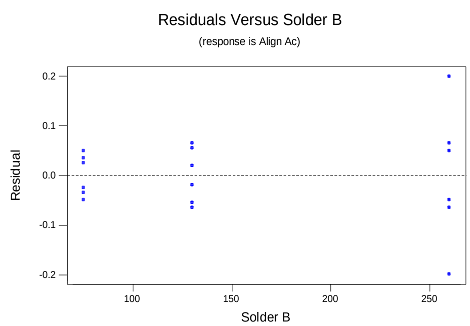


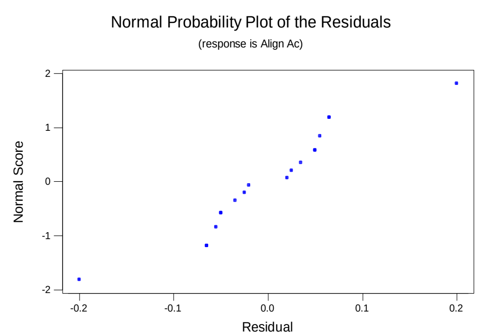
Because the smaller value is preferred, to improve alignment accuracy solder size should be set at its middle level (130μm) while the alignment method used should be method 3. There is not much difference between bump size of 130 μm and 75 μm when method 3 is used.

(c)









The normal probability plot does not suggest a departure from normality. The assumption of constant variance may be of concern. It appears that the variability is lower for the high level of the factor A.

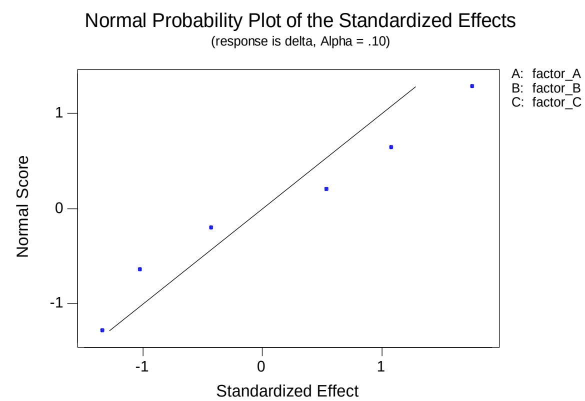
14-79 An article in *Solid State Technology* (1984, Vol. 29, pp. 281–284) described the use of factorial experiments in photolithography, an important step in the process of manufacturing integrated circuits. The variables in this experiment (all at two levels) are prebake temperature (A), prebake time (B), and exposure energy (C), and the response variable is delta line width, the difference between the line on the mask and the printed line on the device. The data are as follows: (1) = –2.30, a = –9.87, b = –18.20, ab = –30.20, c = –23.80, ac = –4.30, bc = –3.80, and abc = –14.70.

(a) Estimate the factor effects.
(b) Use a normal probability plot of the effect estimates to identity factors that may be important.
(c) What model would you recommend for predicting the delta line width response based on the results of this experiment?
(d) Analyze the residuals from this experiment, and comment on model adequacy.

(a) Estimated Effects

Term	Effect
A	-2.74
B	-6.66
C	3.49
A*B	-8.71
A*C	7.04
B*C	11.46
A*B*C	-6.49

(b) None of the factor effects are significant

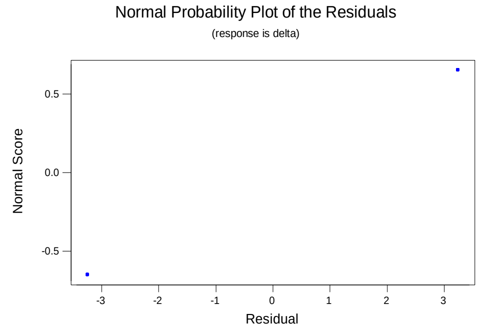


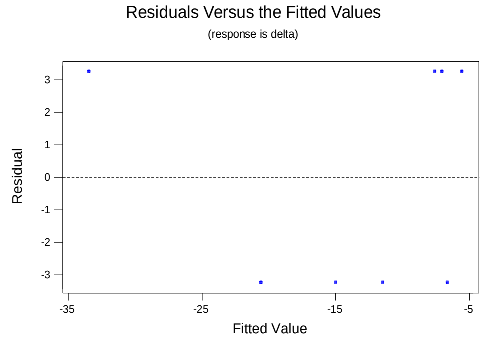
(c)
Analysis of Variance for delta (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	128.07	128.07	42.69	0.51	0.746
2-Way Interactions	3	512.99	512.99	171.00	2.02	0.467
Residual Error	1	84.50	84.50	84.50		
Total	7	725.56				

Based on the analysis of this data, a significant model cannot be built. It appears that more data should be collected or that maybe a factor is missing.

(d) The residual plots shows there are serious problems with the data.





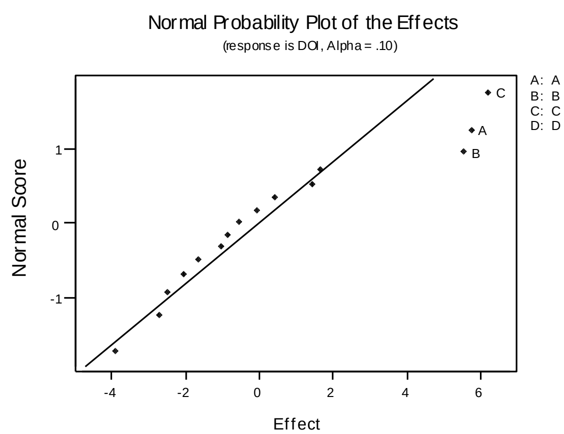
14-80 An article in the *Journal of Coatings Technology* (1988, Vol. 60, pp. 27–32) described a 2^4 factorial design used for studying a silver automobile basecoat. The response variable is distinctness of image (DOI). The variables used in the experiment are

A = Percentage of polyester by weight of polyester/melamine (low value = 50%, high value = 70%)
B = Percentage of cellulose acetate butyrate carboxylate (low value = 15%, high value = 30%)
C = Percentage of aluminum stearate (low value = 1%, high value = 3%)
D = Percentage of acid catalyst (low value = 0.25%, high value = 0.50%)
The responses are (1) = 63.8, *a* = 77.6, *b* = 68.8, *ab* = 76.5, *c* = 72.5, *ac* = 77.2, *bc* = 77.7, *abc* = 84.5, *d* = 60.6, *ad* = 64.9, *bd* = 72.7, *abd* = 73.3, *cd* = 68.0, *acd* = 76.3, *bcd* = 76.0, and *abcd* = 75.9.

(a) Estimate the factor effects.
(b) From a normal probability plot of the effects, identify a tentative model for the data from this experiment.
(c) Using the apparently negligible factors as an estimate of error, test for significance of the factors identified in part (b). Use $\alpha = 0.05$.
(d) What model would you use to describe the process based on this experiment? Interpret the model.
(e) Analyze the residuals from the model in part (d) and comment on your findings.

(a)
```
Estimated Effects and Coefficients for DOI
Term          Effect      Coef
Constant                72.894
A              5.763     2.881
B              5.563     2.781
C              6.238     3.119
D             -3.862    -1.931
A*B           -2.013    -1.006
A*C           -0.837    -0.419
A*D           -2.488    -1.244
B*C           -0.537    -0.269
B*D            1.462     0.731
C*D           -0.063    -0.031
A*B*C          0.437     0.219
A*B*D         -1.013    -0.506
A*C*D          1.663     0.831
B*C*D         -2.687    -1.344
A*B*C*D       -1.612    -0.806
```

(b)



Based on the normal probability plot, it appears that factors A, B, and C are significant.

(c) Conduct an analysis using the main factors A, B, and C and interactions among these variables to check if any are significant.

```
Predictor        Coef        StDev          T         P
Constant       72.894        1.073      67.92     0.000
```

```
A                2.881        1.073        2.68     0.028
B                2.781        1.073        2.59     0.032
C                3.119        1.073        2.91     0.020
AB              -1.006        1.073       -0.94     0.376
AC              -0.419        1.073       -0.39     0.707
BC              -0.269        1.073       -0.25     0.809
ABC              0.219        1.073        0.20     0.844

S = 4.293       R-Sq = 74.6%      R-Sq(adj) = 52.4%
```

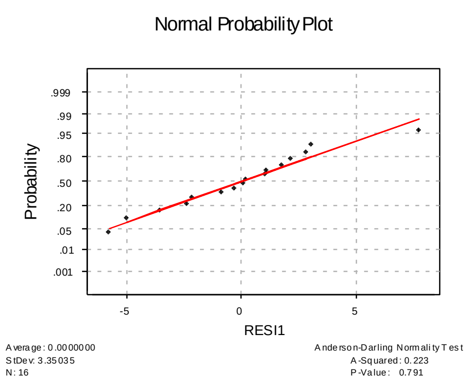
Based on this analysis, only the main effects A, B, and C are significant. Conduct a regression analysis with these important factors.

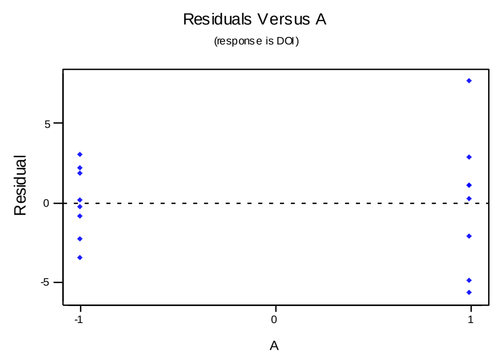
(d)

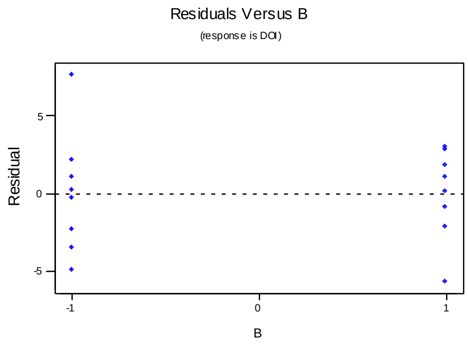
```
The regression equation is
DOI = 72.9 + 2.88 A + 2.78 B + 3.12 C
Predictor        Coef        StDev           T        P
Constant      72.8937       0.9365       77.84    0.000
A              2.8813       0.9365        3.08    0.010
B              2.7813       0.9365        2.97    0.012
C              3.1188       0.9365        3.33    0.006

S = 3.746       R-Sq = 71.0%      R-Sq(adj) = 63.7%

Analysis of Variance
Source        DF          SS          MS         F        P
Regression     3      412.22      137.41      9.79    0.002
Error         12      168.37       14.03
Total         15      580.59

Source        DF      Seq SS
A              1      132.83
B              1      123.77
C              1      155.63
```

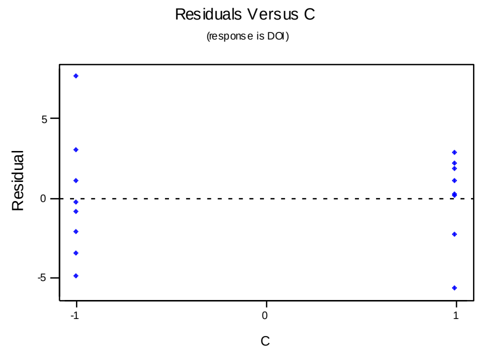
$$\hat{y} = 72.9 + 2.88x_1 + 2.78x_2 + 3.12x_3$$

(e)









The residual plots appear to be adequate.

14-81 An article in the *Journal of Manufacturing Systems* (1991, vol. 10, pp. 32–40) described an experiment to investigate the effect of four factors, *P* = waterjet pressure, *F* = abrasive flow rate, *G* = abrasive grain size, and *V* = jet traverse speed, on the surface roughness of a waterjet cutter. A 2^4 design follows.

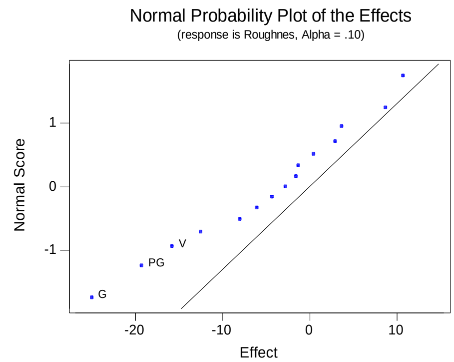
(a) Estimate the factor effects.
(b) Form a tentative model by examining a normal probability plot of the effects.
(c) Is the model in part (b) a reasonable description of the process? Is lack of fit significant? Use α = 0.05.
(d) Interpret the results of this experiment.
(e) Analyze the residuals from this experiment.

	Factors				
Run	*V* (in/min)	*F* (lb/min)	*P* (kpsi)	*G* (Mesh No.)	Surface Roughness (μm)
1	6	2.0	38	80	104
2	2	2.0	38	80	98
3	6	2.0	30	80	103
4	2	2.0	30	80	96
5	6	1.0	38	80	137
6	2	1.0	38	80	112
7	6	1.0	30	80	143
8	2	1.0	30	80	129
9	6	2.0	38	170	88
10	2	2.0	38	170	70
11	6	2.0	30	170	110
12	2	2.0	30	170	110
13	6	1.0	38	170	102
14	2	1.0	38	170	76
15	6	1.0	30	170	98
16	2	1.0	30	170	68

(a)
```
Term          Effect
V             -15.75
F               8.75
P              10.75
G             -25.00
V*F             3.00
V*P            -8.00
V*G            -2.75
F*P            -6.00
F*G             3.75
```

```
P*G          -19.25
V*F*P         -1.25
V*F*G          0.50
V*P*G         -1.50
F*P*G        -12.50
V*F*P*G       -4.25
```

(b)



According to the probability plot, factors V, P, and G and, PG are possibly significant.

Estimated Effects and Coefficients for roughnes (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		102.75	2.986	34.41	0.000
V	-15.75	-7.87	2.986	-2.64	0.046
F	8.75	4.37	2.986	1.46	0.203
P	10.75	5.37	2.986	1.80	0.132
G	-25.00	-12.50	2.986	-4.19	0.009
V*F	3.00	1.50	2.986	0.50	0.637
V*P	-8.00	-4.00	2.986	-1.34	0.238
V*G	-2.75	-1.38	2.986	-0.46	0.665
F*P	-6.00	-3.00	2.986	-1.00	0.361
F*G	3.75	1.88	2.986	0.63	0.558
P*G	-19.25	-9.63	2.986	-3.22	0.023

Analysis of Variance for roughness (coded units)

Analysis of Variance for Roughness (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	4260.7	4260.7	1065.2	7.46	0.024
2-Way Interactions	6	2004.7	2004.7	334.1	2.34	0.184
Residual Error	5	713.5	713.5	142.7		
Total	15	6979.0				

$$\hat{y} = 102.75 - 7.87x_1 + 5.37x_3 - 12.50x_4 - 9.63x_{34}$$

(c) From the analysis, we see that water jet pressure (P), abrasive grain size (G), and jet traverse speed (V) are significant along with the interaction of water jet pressure and abrasive grain size. The model without the interaction is a reasonable model. With the interaction, there is a problem with collinearity. Without the interaction, Factor P is likely not required in the model.

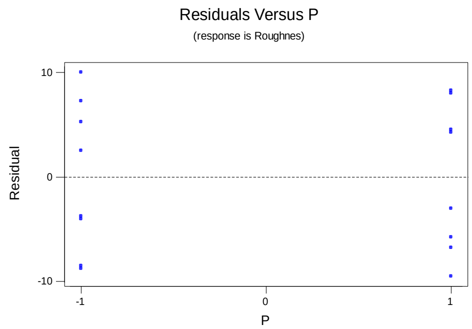
Analysis of Variance for Roughnes, using Adjusted SS for Tests

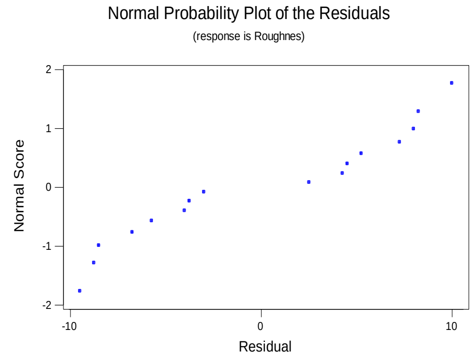
Source	DF	Seq SS	Adj SS	Adj MS	F	P
V	1	992.2	992.2	992.2	3.74	0.077

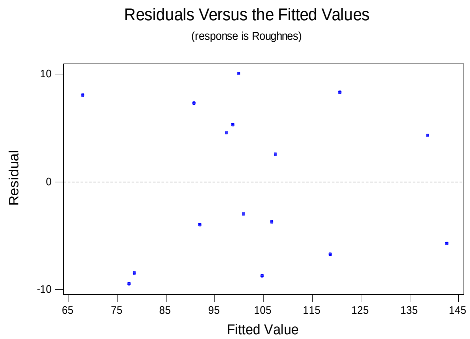
P	1	306.2	306.2	306.2	1.16	0.304
G	1	2500.0	2500.0	2500.0	9.43	0.010
Error	12	3180.5	3180.5	265.0		
Total	15	6979.0				

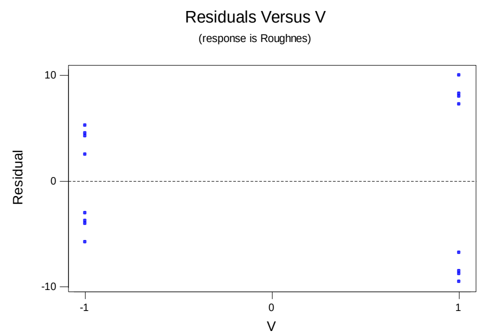
(d) To minimize, abrasive grain size should be at the higher level with jet traverse speed at the lower level.

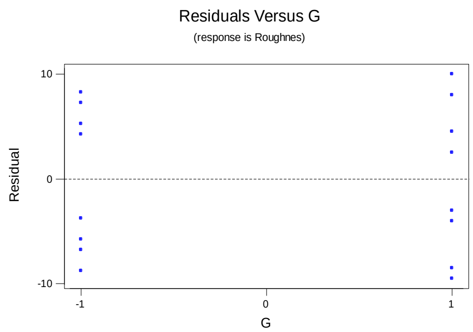
(e) The residual plots appear to indicate the assumption of constant variance may not be met. The assumption of normality appears reasonable.

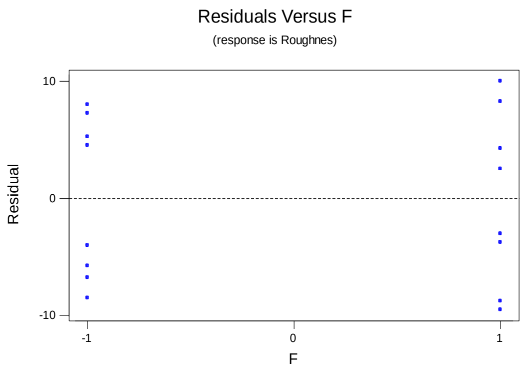












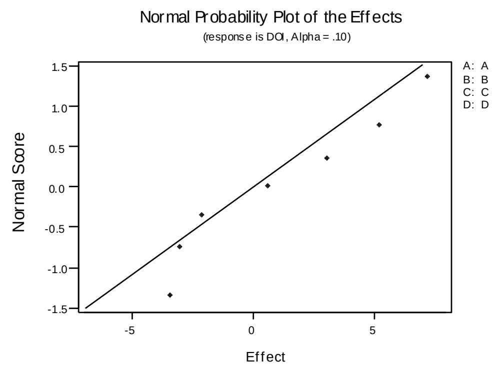
14-82 Construct 2_{IV}^{4-1} design for the problem in Exercise 14-80. Select the data for the eight runs that would have been required for this design. Analyze these runs and compare your conclusions to those obtained in Exercise 14-80 for the full factorial.

One possible 2^{4-1} design is I =ABCD

A	B	C	D	DOI
-1	-1	-1	-1	63.8
1	-1	-1	1	64.9
-1	1	-1	1	72.7
1	1	-1	-1	76.5
-1	-1	1	1	68.0
1	-1	1	-1	77.2
-1	1	1	-1	77.7
1	1	1	1	75.9

Estimated Effects

Term	Effect
A	3.075
B	7.225
C	5.225
D	-3.425
AB	-2.075
AC	0.625
AD	-3.025



It may be useful to conduct an analysis with the main effects only to see which main effect is significant.

```
Term          Effect      Coef  StDev Coef       T      P
Constant                72.088       1.074   67.11  0.000
A              3.075     1.538       1.074    1.43  0.248
B              7.225     3.612       1.074    3.36  0.044
C              5.225     2.612       1.074    2.43  0.093
D             -3.425    -1.712       1.074   -1.59  0.209

Analysis of Variance for DOI
Source                DF      Seq SS      Adj SS      Adj MS       F      P
Main Effects           4      201.38      201.38      50.344    5.45  0.097
Residual Error         3       27.69       27.69       9.231
Total                  7      229.07
```

Based on this analysis, only factor B appears to be significant at the 0.05 level of significance. Comparing this result to that of the previous exercise, factors A and C were not revealed as significant in the smaller design.

14-83 Construct 2_{IV}^{4-1} design for the problem in Exercise 14-81. Select the data for the eight runs that would have been required for this design. Analyze these runs and compare your conclusions to those obtained in Exercise 14-81 for the full factorial.

V	F	P	G	Surface Roughness
-1	-1	-1	-1	129
1	-1	-1	1	98
-1	1	-1	1	110
1	1	-1	-1	103
-1	-1	1	1	76
1	-1	1	-1	137
-1	1	1	-1	98
1	1	1	1	88

```
Estimated Effects and Coefficients for roughness (coded units)

Estimated Effects and Coefficients for surf (coded units)

Term          Effect      Coef     SE Coef       T       P
Constant                104.88       7.950   13.19   0.006
V               3.25      1.63       7.950    0.20   0.857
F             -10.25     -5.12       7.950   -0.64   0.585
P             -10.25     -5.12       7.950   -0.64   0.585
G             -23.75    -11.88       7.950   -1.49   0.274
P*G           -11.75     -5.87       7.950   -0.74   0.537

Alias Structure

I + V*F*P*G
V + F*P*G
F + V*P*G
P + V*F*G
G + V*F*P
V*F + P*G
```

None of the effects are significant for the fractional factorial model. The fractional factorial does not provide enough information to detect significant effects in this experiment.

14-84 Construct a 2_{IV}^{8-4} design in 16 runs. What are the alias relationships in this design?

ABCDEFGH 2^{8-4} Resolution IV

```
Design Generators:  E = BCD  F = ACD  G = ABC  H = ABD

Alias Structure (up to order 4)

I + ABCG + ABDH + ABEF + ACDF + ACEH + ADEG + AFGH + BCDE + BCFH + BDFG + BEGH+
CDGH + CEFG + DEFH

A + BCG + BDH + BEF + CDF + CEH + DEG + FGH
B + ACG + ADH + AEF + CDE + CFH + DFG + EGH
C + ABG + ADF + AEH + BDE + BFH + DGH + EFG
D + ABH + ACF + AEG + BCE + BFG + CGH + EFH
E + ABF + ACH + ADG + BCD + BGH + CFG + DFH
F + ABE + ACD + AGH + BCH + BDG + CEG + DEH
G + ABC + ADE + AFH + BDF + BEH + CDH + CEF
H + ABD + ACE + AFG + BCF + BEG + CDG + DEF
AB + CG + DH + EF + ACDE + ACFH + ADFG + AEGH + BCDF + BCEH + BDEG + BFGH
AC + BG + DF + EH + ABDE + ABFH + ADGH + AEFG + BCDH + BCEF + CDEG + CFGH
AD + BH + CF + EG + ABCE + ABFG + ACGH + AEFH + BCDG + BDEF + CDEH + DFGH
AE + BF + CH + DG + ABCD + ABGH + ACFG + ADFH + BCEG + BDEH + CDEF + EFGH
```

```
AF + BE + CD + GH + ABCH + ABDG + ACEG + ADEH + BCFG + BDFH + CEFH + DEFG
AG + BC + DE + FH + ABDF + ABEH + ACDH + ACEF + BDGH + BEFG + CDFG + CEGH
AH + BD + CE + FG + ABCF + ABEG + ACDG + ADEF + BCGH + BEFH + CDFH + DEGH
```

14-85 Construct a 2^{5-2}_{III} design in eight runs. What are the alias relationships in this design?

ABCDE 2^{5-2} 8 runs

```
Design Generators:  D = AB  E = AC

Alias Structure  I + ABD + ACE + BCDE

A + BD + CE + ABCDE
B + AD + CDE + ABCE
C + AE + BDE + ABCD
D + AB + BCE + ACDE
E + AC + BCD + ABDE
BC + DE + ABE + ACD
BE + CD + ABC + ADE
```

Design

StdOrder	A	B	C	D	E
1	-1	-1	-1	1	1
2	1	-1	-1	-1	-1
3	-1	1	-1	-1	1
4	1	1	-1	1	-1
5	-1	-1	1	1	-1
6	1	-1	1	-1	1
7	-1	1	1	-1	-1
8	1	1	1	1	1

14-86 An article in the *Journal of Quality Technology* (1985, Vol. 17, pp. 198–206) described the use of a replicated fractional factorial to investigate the effect of five factors on the free height of leaf springs used in an automotive application. The factors are A = furnace temperature, B = heating time, C = transfer time, D = hold down time, and E = quench oil temperature. The data are shown in the following table.

(a) What is the generator for this fraction? Write out the alias structure.
(b) Analyze the data. What factors influence mean free height?
(c) Calculate the range of free height for each run. Is there any indication that any of these factors affect variability in free height?
(d) Analyze the residuals from this experiment and comment on your findings.

A	*B*	*C*	*D*	*E*	Free Height		
–	–	–	–	–	7.78	7.78	7.81
+	–	–	+	–	8.15	8.18	7.88
–	+	–	+	–	7.50	7.56	7.50
+	+	–	–	–	7.59	7.56	7.75
–	–	+	+	–	7.54	8.00	7.88
+	–	+	–	–	7.69	8.09	8.06
–	+	+	–	–	7.56	7.52	7.44
+	+	+	+	–	7.56	7.81	7.69
–	–	–	–	+	7.50	7.56	7.50
+	–	–	+	+	7.88	7.88	7.44
–	+	–	+	+	7.50	7.56	7.50
+	+	–	–	+	7.63	7.75	7.56
–	–	+	+	+	7.32	7.44	7.44
+	–	+	–	+	7.56	7.69	7.62
–	+	+	–	+	7.18	7.18	7.25
+	+	+	+	+	7.81	7.50	7.59

(a) The generator for this fraction was I = ABCD

```
I = A*B*C*D
A = B*C*D
B = A*C*D
C = A*B*D
D = A*B*C
E = A*B*C*D*E
A*B = C*D
A*C = B*D
A*D = B*C
A*E = B*C*D*E
B*E = A*C*D*E
C*E = A*B*D*E
D*E = A*B*C*E
A*B*E = C*D*E
A*C*E = B*D*E
A*D*E = B*C*E
```

(b) Estimated Effects and Coefficients for freeheig

Term	Effect	Coef	StDev Coef	T	P
Constant		7.6400	0.01901	401.97	0.000
A	0.2133	0.1067	0.01901	5.61	0.000
B	-0.1925	-0.0963	0.01901	-5.06	0.000
C	-0.0783	-0.0392	0.01901	-2.06	0.048
D	0.0625	0.0313	0.01901	1.64	0.110
E	-0.2100	-0.1050	0.01901	-5.52	0.000
A*B	-0.0008	-0.0004	0.01901	-0.02	0.983
A*C	0.0300	0.0150	0.01901	0.79	0.436
A*D	0.0058	0.0029	0.01901	0.15	0.879
A*E	0.0350	0.0175	0.01901	0.92	0.364
B*E	0.1242	0.0621	0.01901	3.27	0.003
C*E	-0.0617	-0.0308	0.01901	-1.62	0.115
D*E	0.0108	0.0054	0.01901	0.28	0.777
A*B*E	0.0308	0.0154	0.01901	0.81	0.423
A*C*E	0.0483	0.0242	0.01901	1.27	0.213
A*D*E	-0.0308	-0.0154	0.01901	-0.81	0.423

Analysis of Variance for freeheig

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	1.64052	1.64052	0.32810	18.92	0.000
2-Way Interactions	7	0.25797	0.25797	0.03685	2.13	0.069
3-Way Interactions	3	0.05085	0.05085	0.01695	0.98	0.416
Residual Error	32	0.55487	0.55487	0.01734		
Pure Error	32	0.55487	0.55487	0.01734		
Total	47	2.50420				

Based on the analysis, factors A, B, C, and E are significant. The interaction of BE is also significant.

(c)

A	B	C	D	E	Range
-1	-1	-1	-1	-1	0.03
1	-1	-1	1	-1	0.30
-1	1	-1	1	-1	0.06
1	1	-1	-1	-1	0.19
-1	-1	1	1	-1	0.46
1	-1	1	-1	-1	0.40
-1	1	1	-1	-1	0.12
1	1	1	1	-1	0.25
-1	-1	-1	-1	1	0.06
1	-1	-1	1	1	0.44
-1	1	-1	1	1	0.06
1	1	-1	-1	1	0.19

-1	-1	1	1	1	0.12
1	-1	1	-1	1	0.13
-1	1	1	-1	1	0.07
1	1	1	1	1	0.31

Estimated Effects and Coefficients for Range

Term	Effect	Coef	StDev Coef	T	P
Constant		0.19938	0.02714	7.35	0.000
A	0.15375	0.07688	0.02714	2.83	0.018
B	-0.08625	-0.04313	0.02714	-1.59	0.143
C	0.06625	0.03312	0.02714	1.22	0.250
D	0.10125	0.05062	0.02714	1.87	0.092
E	-0.05375	-0.02687	0.02714	-0.99	0.345

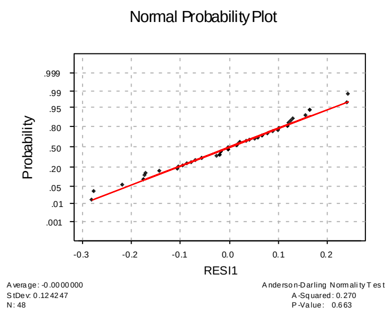
Analysis of Variance for Range

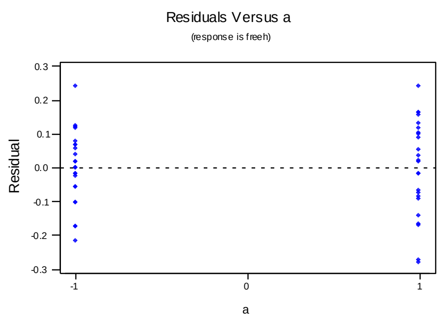
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	0.1944	0.1944	0.03889	3.30	0.051
Residual Error	10	0.1179	0.1179	0.01179		
Total	15	0.3123				

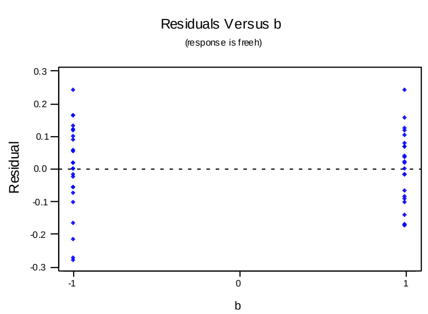
From the analysis, factor A is significant for variability in free height.

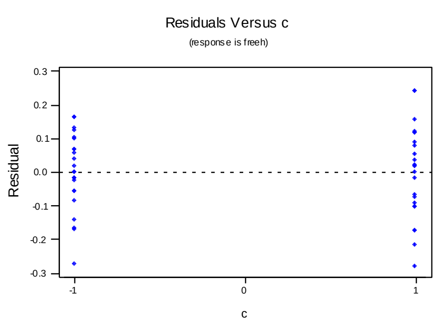
Using the model $\hat{y} = 0.19938 + 0.07688x_1$

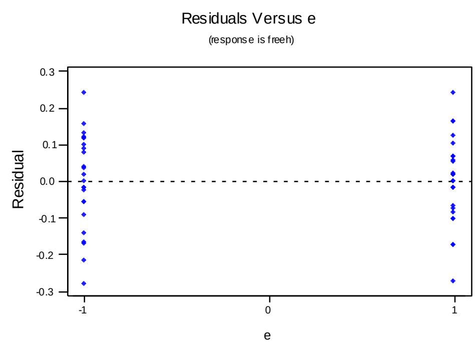
(d)











The residual plots appear to be adequate.

14-87 An article in *Rubber Chemistry and Technology* (Vol. 47, 1974, pp. 825–836) described an experiment to study the effect of several variables on the Mooney viscosity of rubber, including silica filler (parts per hundred) and oil filler (parts per hundred). Data typical of that reported in this experiment are reported in the following table where

$$x_1 = \frac{\text{silica}-60}{15}, x_2 = \frac{\text{oil}-21}{15}$$

(a) What type of experimental design has been used?
(b) Analyze the data and draw appropriate conclusions.

Coded levels		
x_1	x_2	y
–1	–1	13.71
1	–1	14.15
–1	1	12.87
1	1	13.53
–1	–1	13.90
1	–1	14.88
–1	1	12.25
–1	1	13.35

(a) The design used is a 2^2 full factorial with 2 replicates.
(b) Factors x_1 and x_2 are significant. The interaction between x_1 and x_2 is not significant

```
Term           Effect      Coef     SE Coef       T       P
Constant                13.5800      0.1241  109.42   0.000
x1             0.7950    0.3975      0.1241    3.20   0.033
x2            -1.1600   -0.5800      0.1241   -4.67   0.009
x1*x2          0.0850    0.0425      0.1241    0.34   0.749

Analysis of Variance for y (coded units)
Source                  DF     Seq SS     Adj SS     Adj MS      F      P
Main Effects             2    3.95525    3.95525    1.97763  16.05  0.012
2-Way Interactions       1    0.01445    0.01445    0.01445   0.12  0.749
Residual Error           4    0.49290    0.49290    0.12323
  Pure Error             4    0.49290    0.49290    0.12322
Total                    7    4.46260
```

14-88 An article in *Tropical Science* ["Proximate Composition of the Seeds of *Acacia Nilotica* var *Adansonti* (Bagaruwa) and Extraction of Its Protein" (1992, Vol. 32(3), pp. 263–268)] reported on research extracting and concentrating the proteins of the bagaruwa seeds in livestock feeding in Nigeria to eliminate the toxic substances from the seeds. The following are the effects of extraction time and flour to solvent ratio on protein extractability of the bagaruwa seeds in distilled water:

Flour: Solvent Ratio (w/v) (%)	Percentage of Protein Extracted at Time (min)			
	30	60	90	120
3	30.5	45.7	30.5	31.0
	36.9	44.3	29.5	22.1
7	32.9	42.4	28.2	23.5
	37.5	40.9	27.3	34.1
11	29.0	39.5	29.0	29.0
	32.7	43.6	30.5	28.4

All values are means of three determinations.

(a) Test the appropriate hypotheses and draw conclusions using the analysis of variance with $\alpha = 0.5$.
(b) Graphically analyze the interaction.
(c) Analyze the residuals from this experiment.

(a)

1. H0: $\tau_1 = \tau_2 = \tau_3 = 0$

   H1: at least one $\tau_i \neq 0$

2. H0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

   H1: at least one $\beta_j \neq 0$

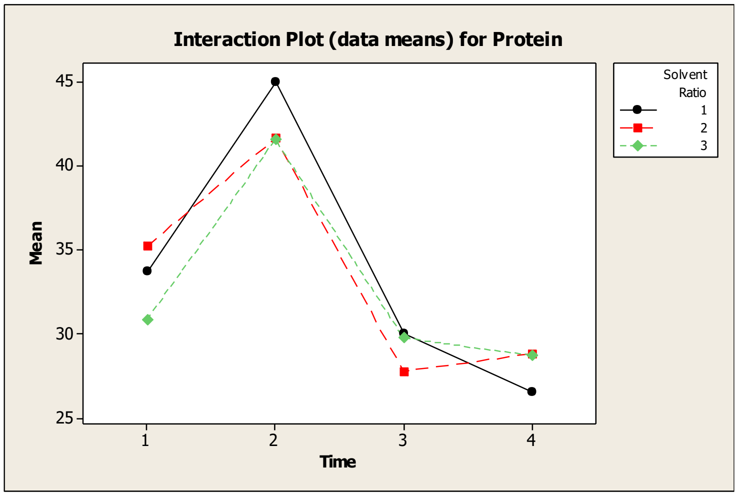
3. H0: $(\tau\beta)_{11} = (\tau\beta)_{12} = (\tau\beta)_{13} = \ldots = (\tau\beta)_{34} = 0$

   H1: at least one $(\tau\beta)_{ij} \neq 0$

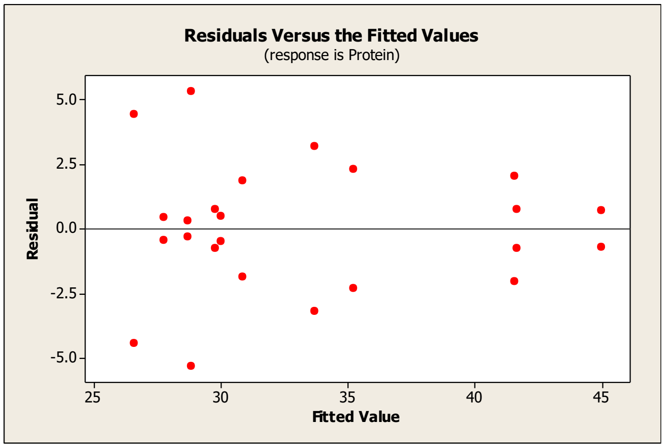
```
Analysis of Variance for Protein, using Adjusted SS for Tests

Source              DF   Seq SS   Adj SS   Adj MS      F      P
Solvent Ratio        2     4.88     4.88     2.44   0.20  0.821
Time                 3   803.93   803.93   267.98  21.96  0.000
Solvent Ratio*Time   6    42.62    42.62     7.10   0.58  0.738
Error               12   146.41   146.41    12.20
Total               23   997.84
```

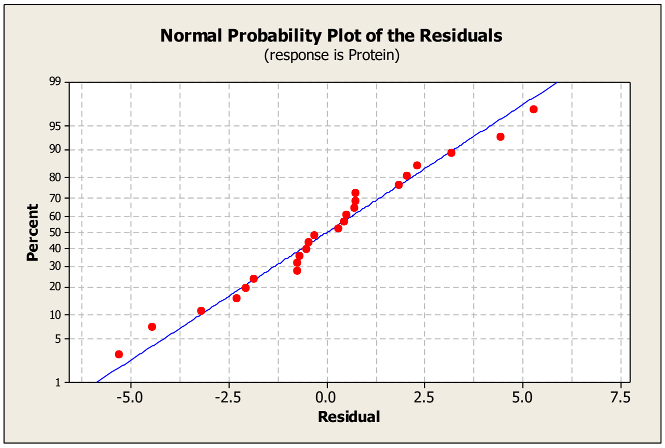
The only the time effect is significant.

(b) The mean percentage of protein extracted of solvent 2 highest at time1, solvent 2 is highest at time 2 and 3, but lowest at time 4. The lines cross, but they are approximately parallel. This supports the ANOVA results that the interaction is not significant.



(c) The plot of the residuals versus the fitted values shows a concern with the assumption of equal variances. The normality assumption appears reasonable.





14-89 An article in *Plant Disease* ["Effect of Nitrogen and Potassium Fertilizer Rates on Severity of Xanthomonas Blight of *Syngonium Podophyllum*" (1989, Vol. 73(12), pp. 972–975)] showed the effect of the variable nitrogen and potassium rates on the growth of "White Butterfly" and the mean percentage of disease. Data representative of that collected in this experiment is provided in the following table.

Nitrogen (mg/pot/wk)	Potassium (mg/pot/wk)		
	30	90	120
50	61.0 61.3	45.5 42.5	59.5 58.2
150	54.5 55.9	53.5 51.9	34.0 35.9
250	42.7 40.4	36.5 37.4	32.5 33.8

(a) State the appropriate hypotheses.
(b) Use the analysis of variance to test these hypotheses with α = 0.05.
(c) Graphically analyze the residuals from this experiment.
(d) Estimate the appropriate variance components.

(a)

1. H_0: $\tau_1 = \tau_2 = \tau_3 = 0$

   H_1: at least one $\tau_i \neq 0$

2. H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

   H_1: at least one $\beta_j \neq 0$

3. H_0: $(\tau\beta)_{11} = (\tau\beta)_{12} = (\tau\beta)_{13} = ... = (\tau\beta)_{33} = 0$

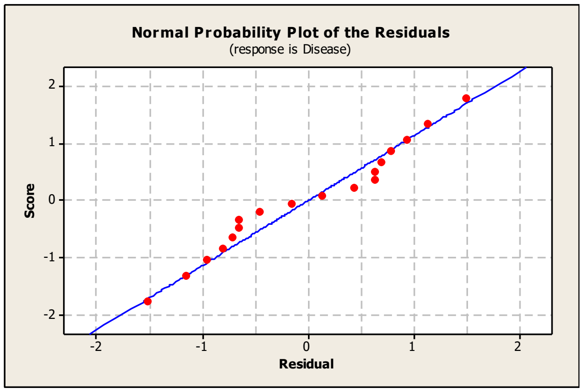
   H_1: at least one $(\tau\beta)_{ij} \neq 0$

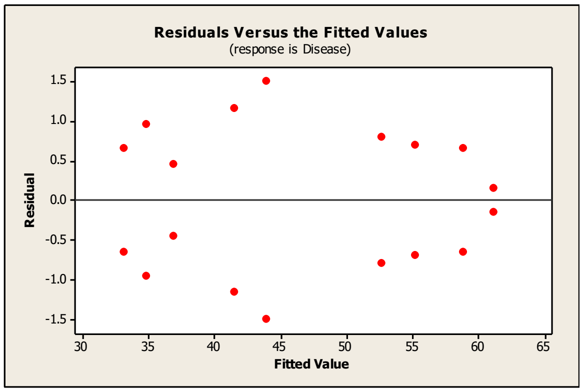
(b)

```
Analysis of Variance for Disease

Source              DF       SS      MS       F      P
Nitrogen             2   924.73  462.37  311.71  0.000
Potassium            2   353.52  176.76  119.17  0.000
Nitrogen*Potassium   4   551.46  137.86   92.94  0.000
Error                9    13.35    1.48
Total               17  1843.06

S = 1.21792   R-Sq = 99.28%   R-Sq(adj) = 98.63%
```

All effects in the model are significant.

(c) The residual plots do not show any violations of the model assumptions.





(d) s = 1.21792 from the ANOVA estimates σ

14-90 An article in *Biotechnology Progress* (2001, Vol. 17, pp. 366–368) reported on an experiment to investigate and optimize the operating conditions of the nisin extraction in aqueous two-phase systems (ATPS). A 2^2 full factorial design with center points was used to verify the most significant factors affecting the nisin recovery. The factor *x*1 was the concentration (% w/w) of PEG 4000 and x_1 was the concentration (% w/w) of Na_2SO_4. See the following table for the range and levels of the variables investigated in this study. Nisin extraction is a ratio representing the concentration of nisin, and this was the response *y*.

Trial	x_1	x_2	y
1	13	11	62.874
2	15	11	76.133
3	13	13	87.467
4	15	13	102.324
5	14	12	76.187
6	14	12	77.523
7	14	12	76.782
8	14	12	77.438
9	14	12	78.742

(a) Compute an ANOVA table for the effects and test for curvature with α = 0.05. Is curvature important in this region of the factors?
(b) Calculate residuals from the linear model and test for adequacy of your model.
(c) In a new region of factor space, a central composite design (CCD) was used to perform second-order optimization. The results are shown in the following table. Fit a second order model to this data and make conclusions.

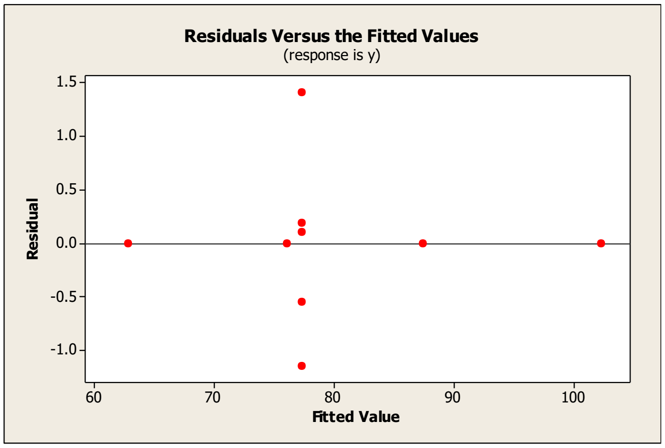
	Coded		Uncoded		
Trial	x_1	x_2	x_1	x_2	y
1	–1	–1	15	14	102.015
2	1	–1	16	14	106.868
3	–1	1	15	16	108.13
4	1	1	16	16	110.176
5	–1.414	0	14.793	15	105.236
6	1.414	0	16.207	15	110.289
7	0	–1.414	15.5	13.586	103.999
8	0	1.414	15.5	16.414	110.171
9	0	0	15.5	15	108.044
10	0	0	15.5	15	109.098
11	0	0	15.5	15	107.824
12	0	0	15.5	15	108.978
13	0	0	15.5	15	109.169

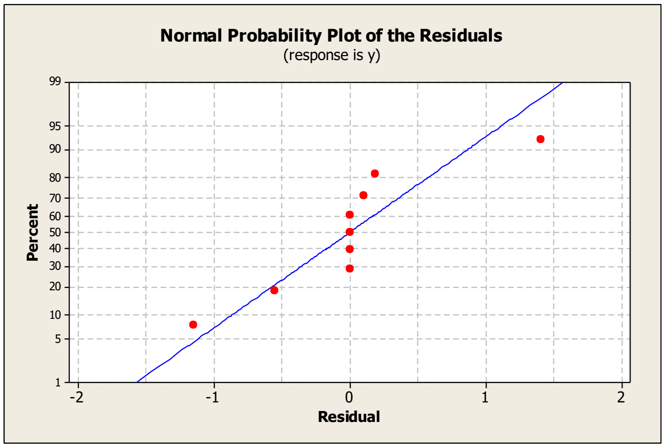
```
Analysis of Variance for y (coded units)

Source                DF   Seq SS   Adj SS   Adj MS       F      P
Main Effects           2  842.381  842.381  421.191  461.67  0.000
2-Way Interactions     1    0.638    0.638    0.638    0.70  0.450
  Curvature            1   52.598   52.598   52.598   57.65  0.002
Residual Error         4    3.649    3.649    0.912
  Pure Error           4    3.649    3.649    0.912
Total                  8  899.267
```

Yes, the curvature is important in this region of the factors because the P-value =0.002 is smaller than $\alpha = 0.05$.

(b) Residual

Trial	X1	X2	y	Residual
1	0	0	76.187	-1.1474
2	-1	-1	62.874	-0.0000
3	0	0	77.523	0.1886
4	1	-1	76.133	0.0000
5	1	1	102.324	0.0000
6	0	0	76.782	-0.5524
7	0	0	77.438	0.1036
8	-1	1	87.467	0.0000
9	0	0	78.742	1.4076





The linear model does not provide a good fit to this data.

(c)

```
Estimated Regression Coefficients for y

Term          Coef  SE Coef        T      P
Constant   108.623   0.2660  408.367  0.000
x1           1.756   0.2103    8.349  0.000
x2           2.269   0.2103   10.790  0.000
x1*x1       -0.587   0.2255   -2.602  0.035
x2*x2       -0.925   0.2255   -4.104  0.005
x1*x2       -0.702   0.2974   -2.360  0.050

S = 0.5948   R-Sq = 96.8%   R-Sq(adj) = 94.5%

Analysis of Variance for y
Source          DF   Seq SS   Adj SS   Adj MS      F      P
Regression       5  75.3066  75.3066  15.0613  42.57  0.000
  Linear         2  65.8426  65.8426  32.9213  93.06  0.000
  Square         2   7.4942   7.4942   3.7471  10.59  0.008
  Interaction    1   1.9698   1.9698   1.9698   5.57  0.050
Residual Error   7   2.4763   2.4763   0.3538
  Lack-of-Fit    3   0.8529   0.8529   0.2843   0.70  0.599
  Pure Error     4   1.6234   1.6234   0.4059
Total           12  77.7829
```

The model is $\hat{y} = 108.623 + 1.756x_1 + 2.269x_2 - 0.587x_1^2 - 0.925x_2^2 - 0.702x_1x_2$

This model is a better fit than the model from part (a).

14-91 An article in the *Journal of Applied Electrochemistry* (May 2008, Vol. 38(5), pp. 583–590) presented a 27–3 fractional factorial design to perform optimization of polybenzimidazolebased membrane electrode assemblies for H_2/O_2 fuel cells. The design and data are shown in the following table.

Runs	*A*	*B*	*C*	*D*	*E*	*F*	*G*	Current Density (CD mA cm²)
1	−1	−1	−1	−1	−1	−1	−1	160
2	+1	−1	−1	−1	+1	+1	+1	20
3	−1	+1	−1	−1	+1	+1	−1	80
4	+1	+1	−1	−1	−1	−1	+1	317
5	−1	−1	+1	−1	+1	−1	+1	19
6	+1	−1	+1	−1	−1	+1	−1	4
7	−1	+1	+1	−1	−1	+1	+1	20
8	+1	+1	+1	−1	+1	−1	−1	88
9	−1	−1	−1	+1	−1	+1	+1	1100
10	+1	−1	−1	+1	+1	−1	−1	12
11	−1	+1	−1	+1	+1	−1	+1	552
12	+1	+1	−1	+1	−1	+1	−1	880
13	−1	−1	+1	+1	+1	+1	−1	16
14	+1	−1	+1	+1	−1	−1	+1	20
15	−1	+1	+1	+1	−1	−1	−1	8
16	+1	+1	+1	+1	+1	+1	+1	15

The factors and levels are shown in the following table.

	Factor	−1	+1
A	Amount of binder in the catalyst layer	0.2 mg cm²	1 mg cm²
B	Electrocatalyst loading	0.1 mg cm²	1 mg cm²
C	Amount of carbon in the gas diffusion layer	2 mg cm²	4.5 mg cm²
D	Hot compaction time	1 min	10 min
E	Compaction temperature	100°C	150°C
F	Hot compaction load	0.04 ton cm²	0.2 ton cm²
G	Amount of PTFE in the gas diffusion layer	0.1 mg cm²	1 mg cm²

(a) Write down the alias relationships.
(b) Estimate the main effects.
(c) Prepare a normal probability plot for the effects and interpret the results.
(d) Calculate the sum of squares for the alias set that contains the ABG interaction from the corresponding effect estimate.

(a) Generators are E = ABC, F = ABD, and G = ACD

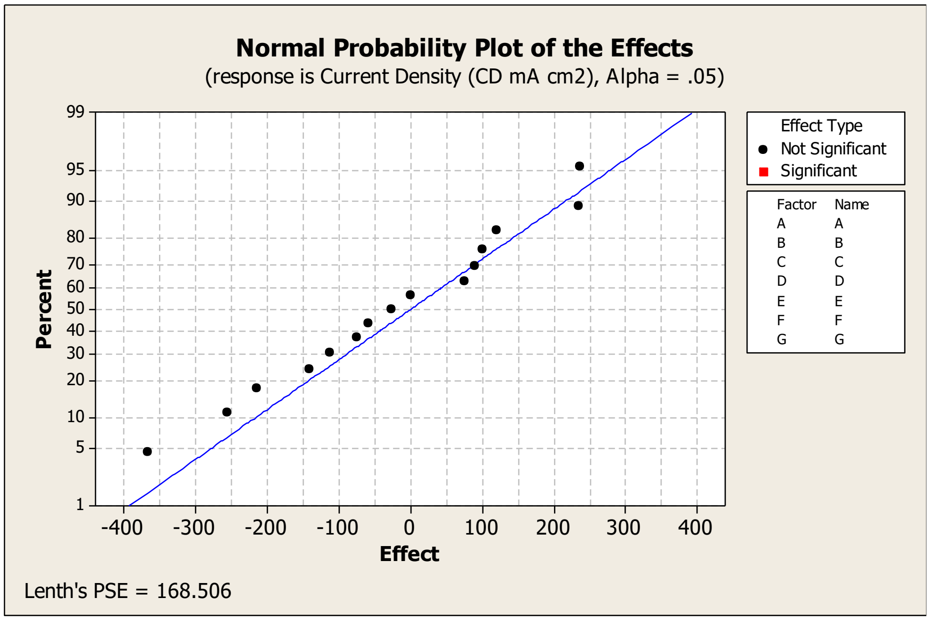
I = ABCE = ABDF = CDEF = ACDG = BDEG = BCFG = AEFG

```
Alias Structure (up to order 3)
I
A + B*C*E + B*D*F + C*D*G + E*F*G
B + A*C*E + A*D*F + C*F*G + D*E*G
C + A*B*E + A*D*G + B*F*G + D*E*F
D + A*B*F + A*C*G + B*E*G + C*E*F
E + A*B*C + A*F*G + B*D*G + C*D*F
F + A*B*D + A*E*G + B*C*G + C*D*E
G + A*C*D + A*E*F + B*C*F + B*D*E
A*B + C*E + D*F
A*C + B*E + D*G
A*D + B*F + C*G
A*E + B*C + F*G
A*F + B*D + E*G
A*G + C*D + E*F
B*G + C*F + D*E
A*B*G + A*C*F + A*D*E + B*C*D + B*E*F + C*E*G + D*F*G
```

(b)
Factorial Fit: Current Density (CD mA cm2) versus A, B, C, D, E, F, G

```
Estimated Effects and Coefficients for Current Density (CD mA cm2) (coded
     units)
Term      Effect    Coef
Constant           206.9
A          -74.9   -37.5
B           76.1    38.0
C         -366.4  -183.2
D          236.9   118.5
E         -213.4  -106.7
F          119.9    60.0
G          101.9    51.0
A*B        234.8   117.4
A*C         90.8    45.4
A*D       -112.3   -56.2
A*E        -58.2   -29.1
A*F          0.7     0.3
A*G       -254.8  -127.4
B*G       -139.8   -69.9
A*B*G      -25.1   -12.5
```

(c) Although the effects C, D, E, F, G, AD, AG, and BG are large, these effects are not indicated as significant in the normal probability plot of the effects.



(d) From part (b), the effect of ABG interaction = -25.1. The contrast of the ABG interactions

= (Effect of ABG interactions x 2^{8-4})/2 = -25.1 x 2^3 = -200.8

The sum of square for the ABG interaction

= Contrast of ABG interactions^2 / 2^{8-4} = 2520 (= 2518 with more precision from computer software)

14-92 An article in *Biotechnology Progress* (December 2002, Vol. 18(6), pp. 1170–1175) presented a 2^{7-3} fractional factorial to evaluate factors promoting astaxanthin production. The data are shown in the following table.

Runs	*A*	*B*	*C*	*D*	*E*	*F*	*G*	Weight Content (mg/g)	Cellular Content (pg/cell)
1	–1	–1	–1	1	1	1	–1	4.2	10.8
2	1	–1	–1	–1	–1	1	1	4.4	24.9
3	–1	1	–1	–1	1	–1	1	7.8	27.3
4	1	1	–1	1	–1	–1	–1	14.9	36.3
5	–1	–1	1	1	–1	–1	1	25.3	112.6
6	1	–1	1	–1	1	–1	–1	26.7	159.3
7	–1	1	1	–1	–1	1	–1	23.9	145.2
8	1	1	1	1	1	1	1	21.9	243.2
9	1	1	1	–1	–1	–1	1	24.3	72.1
10	–1	1	1	1	1	–1	–1	20.5	112.2
11	1	–1	1	1	–1	1	–1	10.8	22.5
12	–1	–1	1	–1	1	1	1	20.8	149.7
13	1	1	–1	–1	1	1	–1	13.5	140.1
14	–1	1	–1	1	–1	1	1	10.3	47.3
15	1	–1	–1	1	1	–1	1	23.0	153.2
16	–1	–1	–1	–1	–1	–1	–1	12.1	35.2

The factors and levels are shown in the following table.

	Factor	−1	+1
A	Nitrogen concentration (mM)	4.06	0
B	Phosphorus concentration (mM)	0.21	0
C	Photon flux density (μE m^{-2} s^{-2})	100	500
D	Magnesium concentration (mM)	1	0
E	Acetate concentration (mM)	0	15
F	Ferrous concentration (mM)	0	0.45
G	NaCl concentration (mM)	OHM	25
OHM: Optimal Haematococcus Medium			

(a) Write down the complete defining relation and the aliases from the design.
(b) Estimate the main effects.
(c) Plot the effect estimates on normal probability paper and interpret the results.

(a) Generators are E = BCD, F = ACD, and G = ABC

I = BCDE = ACDF = ABEF = ABCG = ADEG = BDFG = CEFG

```
Alias Information for Terms in the Model.

I + A*C*D*F + A*B*E*F + A*B*C*G + A*D*E*G + B*C*D*E + B*D*F*G + C*E*F*G
A + B*E*F + B*C*G + C*D*F + D*E*G + A*B*C*D*E + A*B*D*F*G + A*C*E*F*G
B + A*E*F + A*C*G + C*D*E + D*F*G + A*B*C*D*F + A*B*D*E*G + B*C*E*F*G
C + A*D*F + A*B*G + B*D*E + E*F*G + A*B*C*E*F + A*C*D*E*G + B*C*D*F*G
D + A*C*F + A*E*G + B*C*E + B*F*G + A*B*D*E*F + A*B*C*D*G + C*D*E*F*G
E + A*B*F + A*D*G + B*C*D + C*F*G + A*C*D*E*F + A*B*C*E*G + B*D*E*F*G
F + A*C*D + A*B*E + B*D*G + C*E*G + A*B*C*F*G + A*D*E*F*G + B*C*D*E*F
G + A*B*C + A*D*E + B*D*F + C*E*F + A*C*D*F*G + A*B*E*F*G + B*C*D*E*G
A*B + C*G + E*F + A*C*D*E + A*D*F*G + B*C*D*F + B*D*E*G + A*B*C*E*F*G
A*C + B*G + D*F + A*B*D*E + A*E*F*G + B*C*E*F + C*D*E*G + A*B*C*D*F*G
A*D + C*F + E*G + A*B*C*E + A*B*F*G + B*D*E*F + B*C*D*G + A*C*D*E*F*G
A*E + B*F + D*G + A*B*C*D + A*C*F*G + B*C*E*G + C*D*E*F + A*B*D*E*F*G
A*F + B*E + C*D + A*B*D*G + A*C*E*G + B*C*F*G + D*E*F*G + A*B*C*D*E*F
A*G + B*C + D*E + A*B*D*F + A*C*E*F + B*E*F*G + C*D*F*G + A*B*C*D*E*G
B*D + C*E + F*G + A*B*C*F + A*D*E*F + A*C*D*G + A*B*E*G + B*C*D*E*F*G
A*B*D + A*C*E + A*F*G + B*C*F + B*E*G + C*D*G + D*E*F + A*B*C*D*E*F*G
```

(b)
Factorial Fit: Weight versus A, B, C, D, E, F, G

```
Estimated Effects and Coefficients for Weight (coded units)

Term       Effect    Coef
Constant           16.525
A           1.825   0.913
B           1.225   0.612
C          10.500   5.250
D          -0.325  -0.162
E           1.400   0.700
F           1.550   0.775
G          -5.600  -2.800
A*B         1.200   0.600
A*C        -3.525  -1.762
A*D         0.750   0.375
A*E         0.525   0.263
```

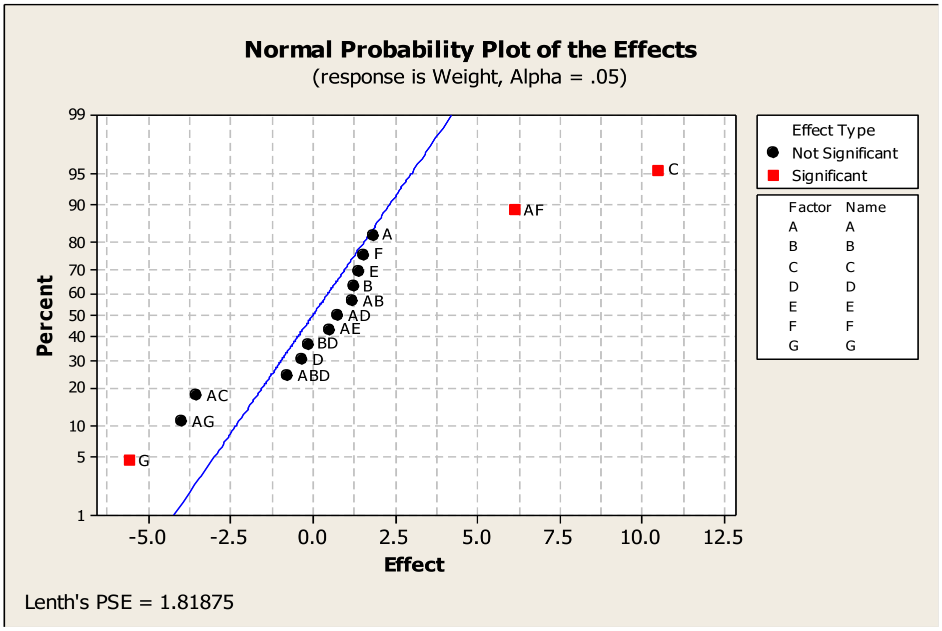
```
A*F        6.125   3.063
A*G       -3.975  -1.987
B*D       -0.150  -0.075
A*B*D     -0.775  -0.387
```

Factorial Fit: Cellular versus A, B, C, D, E, F, G

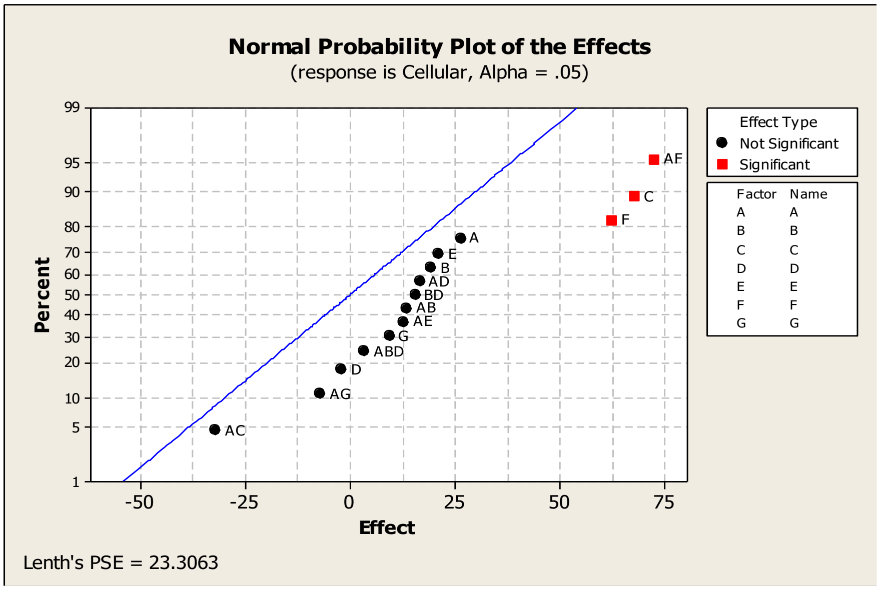
```
Estimated Effects and Coefficients for Cellular (coded units)

Term       Effect    Coef
Constant            93.24
A           26.41   13.21
B           19.44    9.72
C           67.71   33.86
D           -1.96   -0.98
E           21.09   10.54
F           62.46   31.23
G            9.44    4.72
A*B         13.51    6.76
A*C        -32.06  -16.03
A*D         16.66    8.33
A*E         12.71    6.36
A*F         72.54   36.27
A*G         -6.99   -3.49
B*D         15.54    7.77
A*B*D        3.41    1.71
```

(c)

Effects Plot for WeightContent



For weight, the effects labeled as AF, C, and G are marked as significant. Also, AC and AG might be considered important. These effect labels actually represent alias sets and need to be interpreted with the alias table shown above.



For cellular content, the effects labeled as AF, C, and F are marked as significant and AC are also indicated as potentially important. These effect labels actually represent alias sets and need to be interpreted with the alias table shown above.

14-93 The rework time required for a machine was found to depend on the speed at which the machine was run (*A*), the lubricant used while working (*B*), and the hardness of the metal used in the machine (*C*). Two levels of each factor were chosen and a single replicate of a 2^3 experiment was run. The rework time data (in hours) are shown in the following table.

Treatment Combination	Time (in hours)
(1)	27
a	34
b	38
ab	59
c	44
ac	40
bc	63
abc	37

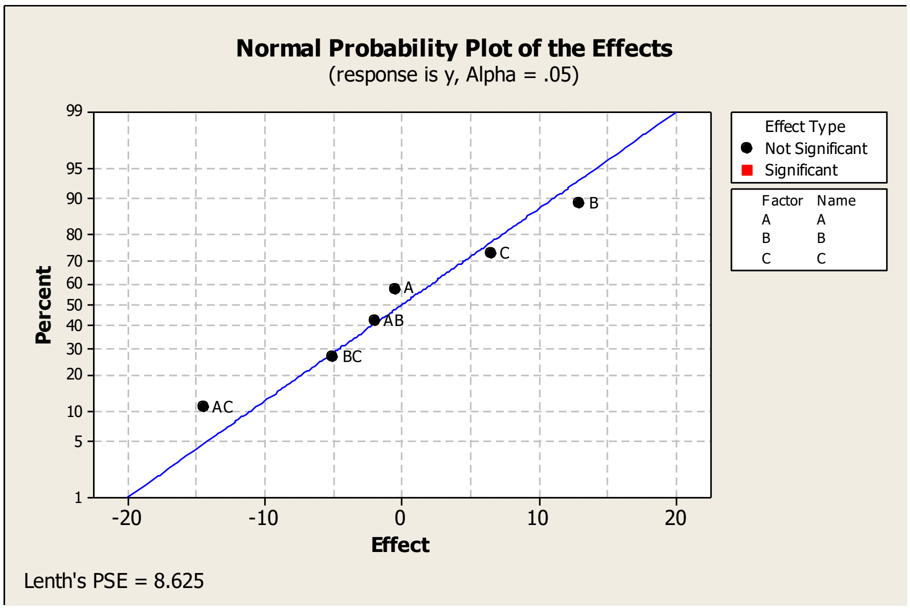
(a) These treatments cannot all be run under the same conditions. Set up a design to run these observations in two blocks of four observations each, with *ABC* confounded with blocks.
(b) Analyze the data.

(a)

Std Order	Run Order	Center Pt	Blocks	A	B	C	y
1	1	1	1	-1	-1	-1	27
2	2	1	1	1	1	-1	59
3	3	1	1	1	-1	1	40
4	4	1	1	-1	1	1	63
5	5	1	2	1	-1	-1	34

6	6	1	2	-1	1	-1	38
7	7	1	2	-1	-1	1	44
8	8	1	2	1	1	1	37

(b) Computer software does not detect any significant effects, but the effects for B, AC, and possibly C are large.



Factorial Fit: Time (in hours) versus Block, A, B, C

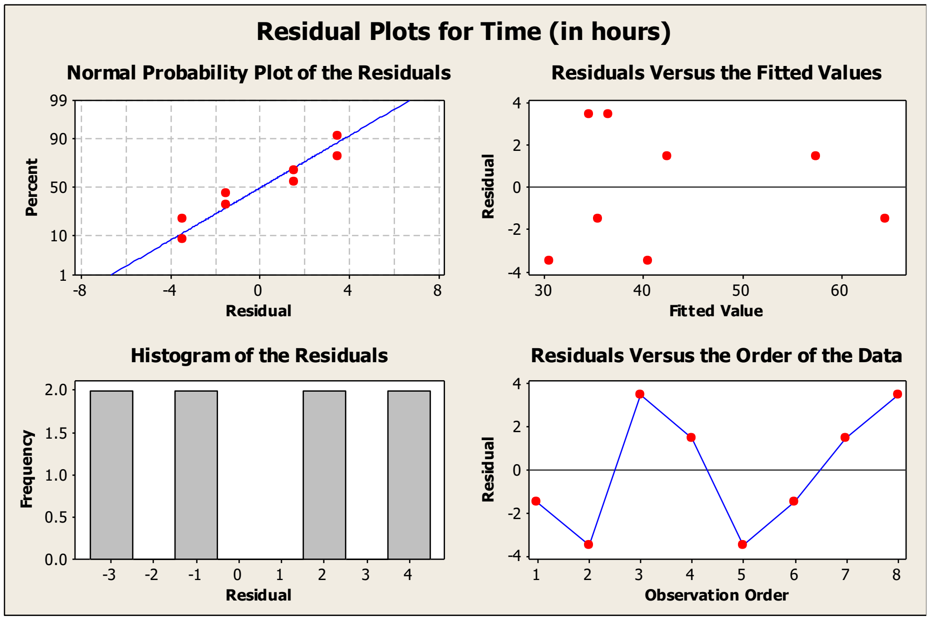
```
Estimated Effects and Coefficients for Time (in hours) (coded units)

Term        Effect     Coef
Constant             42.750
Block                 4.500
A           -0.500   -0.250
B           13.000    6.500
C            6.500    3.250
A*B         -2.000   -1.000
A*C        -14.500   -7.250
B*C         -5.000   -2.500
```

If a hierarchical model is applied, the main effect A is added to the model. The computer results do not detect any significant effects at $\alpha = 0.05$. However, effect B and AC are significant at $\alpha = 0.1$. Residuals plots do not indicate any serious model failures. There is some increased variability at the lower fitted values.

Factorial Fit: Time (in hours) versus Block, A, B, C

```
Estimated Effects and Coefficients for Time (in hours) (coded units)

Term        Effect     Coef  SE Coef      T      P
Constant             42.750    1.904  22.45  0.002
Block                 4.500    1.904   2.36  0.142
A           -0.500   -0.250    1.904  -0.13  0.908
B           13.000    6.500    1.904   3.41  0.076
C            6.500    3.250    1.904   1.71  0.230
A*C        -14.500   -7.250    1.904  -3.81  0.063
```

```
S = 5.38516   R-Sq = 94.55%   R-Sq(adj) = 80.91%

Analysis of Variance for Time (in hours) (coded units)
Source               DF   Seq SS   Adj SS   Adj MS      F      P
Blocks                1   162.00   162.00   162.00   5.59  0.142
Main Effects          3   423.00   423.00   141.00   4.86  0.175
2-Way Interactions    1   420.50   420.50   420.50  14.50  0.063
Residual Error        2    58.00    58.00    29.00
Total                 7  1063.50
```



14-94 Consider the following results from a two-factor experiment with two levels for factor *A* and three levels for factor *B*. Each treatment has three replicates.

A	*B*	Mean	StDev
1	1	21.33333	6.027714
1	2	20	7.549834
1	3	32.66667	3.511885
2	1	31	6.244998
2	2	33	6.557439
2	3	23	10

(a) Calculate the sum of squares for each factor and the interaction.
(b) Calculate the sum of squares total and error.
(c) Complete an ANOVA table with *F*-statistics.

(a)

A	B	Mean	StDev	Sum
1	1	21.3333	6.027714	64
1	2	20	7.549834	60
1	3	32.6667	3.511885	98
2	1	31	6.244998	93
2	2	33	6.557439	99

2	3	23	10	69

	Factor B			
Factor A	1	2	3	$y_{i.}$
1	64	60	98	222
2	93	99	69	261
$y_{.j}$	157	159	167	483 = $y_{..}$

$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn} = \frac{1}{(3)(3)}[222^2 + 261^2] - \frac{483^2}{18} = 84.5$

$SS_B = \frac{1}{an}\sum_{i=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn} = \frac{1}{(2)(3)}[157^2 + 159^2 + 167^2] - \frac{483^2}{18} = 9.3333$

$SS_{Interaction} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_A - SS_B$

$= \frac{1}{(3)}[64^2 + 60^2 + 98^2 + 93^2 + 99^2 + 69^2] - \frac{483^2}{18} - 84.5 - 9.3333 = 449.3333$

(b) `stDev` $= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

For A =1, B = 1, `stDev = 6.027714` $= \sqrt{\frac{\sum_{i=1}^{3}(x_i - 21.3333)^2}{2}}$

$$\sum_{i=1}^{3}(x_i - 21.3333)^2 = 6.027714^2 \times 2 = 72.666672$$

$$(x_1 - 21.3333)^2 + (x_2 - 21.3333)^2 + (x_3 - 21.3333)^2 = 72.66672$$

$$(x_1^2 + x_2^2 + x_3^2) - (2 \times 21.3333)(x_1 + x_2 + x_3) + (3 \times 21.3333^2) = 72.66672$$

$$(x_1^2 + x_2^2 + x_3^2) = 72.66672 + (2 \times 21.3333)(64) - (3 \times 21.3333^2) = 1438$$

A	B	Mean	StDev	sum	sum of (x - xbar)^2	$x_1^2+x_2^2+x_3^2$
1	1	21.3333	6.027714	64	72.666672	1438
1	2	20	7.549834	60	113.999987	1314
1	3	32.6667	3.511885	98	24.666673	3226
2	1	31	6.244998	93	78.000000	2961
2	2	33	6.557439	99	86.000012	3353
2	3	23	10	69	200	1787

$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn} = [1438 + 1314 + 3226 + 2961 + 3353 + 1787] - \frac{483^2}{18} = 1119$

$SS_{Error} = SS_{Total} - SS_A - SS_B - SS_{Interaction} = 1119 - 84.5 - 9.3333 - 449.3333 = 575$

(c) Total trials = 6 treatments * 3 replicates = 18 trials.

The ANOVA table

Source	DF	SS	MS	F	P-value
A	1	84.5	84.5	1.7624	0.209
B	2	9.3333	4.6667	0.0973	0.908
AB	2	449.3333	224.6667	4.686	0.031
Error	12	575	47.9446		
Total	17	1119			

14-95 Consider the following ANOVA table from a two-factor factorial experiment.

Two-way ANOVA: y Versus A, B

Source	DF	SS	MS	F	P
A	3	1213770	404590	?	0.341
B	2	?	17335441	58.30	0.000
Error	?	1784195	?		
Total	11	37668847			

(a) How many levels of each factor were used in the experiment?
(b) How many replicates were used?
(c) What assumption is made in order to obtain an estimate of error?
(d) Calculate the missing entries (denoted with "?") in the ANOVA table.

(a) Factor A has = 3 + 1 = 4 levels. Factor B has 2 + 1 = 3 levels.

(b) The total degrees of freedom =11 which implies the total runs = 12. Therefore, one replicate was used.

(c) The two-factor interaction term (AB) is not significant.

(d) Degree of freedom of error = 11 – 3 – 2 = 6

$$MS(B) = \frac{SS_B}{df_B} = \frac{SS_B}{2} = 17335441$$, then SS_B = 34670882

$$MS_E = \frac{SS_E}{df_E} = \frac{1784195}{6} = 297365.83$$

$$F_A = \frac{MS_A}{MS_E} = \frac{404590}{297365.83} = 1.3601$$

14-96 An article in *Process Biochemistry* (Dec. 1996, Vol. 31(8), pp. 773–785) presented a 2^{7-3} fractional factorial to perform optimization of manganese dioxide bioleaching media. The data are shown in the following table.

Runs	*A*	*B*	*C*	*D*	*E*	*F*	*G*	Manganese Extraction Yield (%)
1	–1	–1	–1	–1	–1	–1	–1	99.0
2	1	–1	–1	–1	1	–1	1	97.4
3	–1	1	–1	–1	1	1	1	97.7
4	1	1	–1	–1	–1	1	–1	90.0
5	–1	–1	1	–1	1	1	–1	100.0
6	1	–1	1	–1	–1	1	1	98.0
7	–1	1	1	–1	–1	–1	1	90.0
8	1	1	1	–1	1	–1	–1	93.5
9	–1	–1	–1	1	–1	1	1	100.0
10	1	–1	–1	1	1	1	–1	98.6
11	–1	1	–1	1	1	–1	–1	97.1
12	1	1	–1	1	–1	–1	1	92.4
13	–1	–1	1	1	1	–1	1	93.0
14	1	–1	1	1	–1	–1	–1	95.0
15	–1	1	1	1	–1	1	–1	97.0
16	1	1	1	1	1	1	1	98.0

The factors and levels are shown in the following table.

	Factor	–1	+1
A	Mineral concentration (%)	10	20
B	Molasses (g/liter)	100	200
C	NH_4NO_3 (g/liter)	1.25	2.50
D	KH_2PO_4 (g/liter)	0.75	1.50
E	$MgSO_4$ (g/liter)	0.5	1.00
F	Yeast extract (g/liter)	0.20	0.50
G	$NaHCO_3$ (g/liter)	2.00	4.00

(a) Write down the complete defining relation and the aliases from the design.
(b) Estimate the main effects.
(c) Plot the effect estimates on normal probability paper and interpret the results
(d) Conduct a residual analysis.

(a) Generators are E = ABC, F = BCD, and G = ABD. Note that the generator for factor G differs from the Minitab default.

I = ABCE = ABDG = CDEG = ACFG = BEFG = BCDF =ADEF

```
Alias Structure (up to order 3)
I
```

```
A + B*C*E + B*D*G + C*F*G + D*E*F
B + A*C*E + A*D*G + C*D*F + E*F*G
C + A*B*E + A*F*G + B*D*F + D*E*G
D + A*B*G + A*E*F + B*C*F + C*E*G
E + A*B*C + A*D*F + B*F*G + C*D*G
F + A*C*G + A*D*E + B*C*D + B*E*G
G + A*B*D + A*C*F + B*E*F + C*D*E
A*B + C*E + D*G
A*C + B*E + F*G
A*D + B*G + E*F
A*E + B*C + D*F
A*F + C*G + D*E
A*G + B*D + C*F
B*F + C*D + E*G
```

(b)
Factorial Fit: Yield(%) versus A, B, C, D, E, F, G

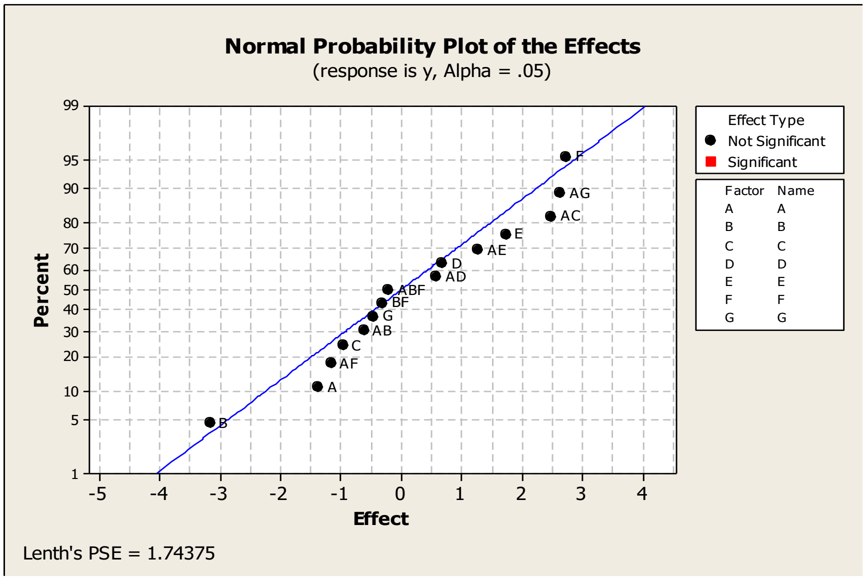
Estimated Effects and Coefficients for Yield(%) (coded units)

Term	Effect	Coef
Constant		96.044
A	-1.362	-0.681
B	-3.162	-1.581
C	-0.962	-0.481
D	0.687	0.344
E	1.738	0.869
F	2.737	1.369
G	-0.462	-0.231

(c)
Factorial Fit: y versus A, B, C, D, E, F, G

Estimated Effects and Coefficients for y (coded units)

Term	Effect	Coef
Constant		96.044
A	-1.362	-0.681
B	-3.162	-1.581
C	-0.962	-0.481
D	0.687	0.344
E	1.738	0.869
F	2.737	1.369
G	-0.462	-0.231
A*B	-0.612	-0.306
A*C	2.487	1.244
A*D	0.587	0.294
A*E	1.287	0.644
A*F	-1.163	-0.581
A*G	2.638	1.319
B*F	-0.312	-0.156
A*B*F	-0.213	-0.106



The computer effects plot does not indicate any significant effects. However, effects B, F, AG, AC are large (in absolute value). A model with the smaller effects A, AF, C, AB, G, BF, ABF pooled into error could be used to test the other effects. These effects are labels for the alias sets in the table above and the aliases need to be used to interpret these results.

(d) $MS_{\text{Two-way Interaction}} = \dfrac{SS_{Two-way\ Interaction}}{df_{Two-way\ Interaction}} = \dfrac{67.884}{7} = 9.698$

$SS_{\text{Residual Error}} = SS_{\text{Total}} - SS_{\text{Main Effects}} - SS_{\text{Two-way Interactioin}} = 163.999 – 95.934- 67.884 = 0.181$

$MS_{\text{Residual Error}} = \dfrac{SS_{Residual Error}}{df_{Residual Error}} = \dfrac{0.181}{1} = 0.181$

F-test = $\dfrac{MS_{Main\ Effects}}{MS_{Residual Error}} = \dfrac{13.7049}{0.181} = 75.72$

P-value = 0.088

14-97 An article in *European Food Research and Technology* [“Factorial Design Optimisation of Grape (Vitis vinifera) Seed Polyphenol Extraction” (2009, Vol. 229(5), pp. 731–742)] used a central composite design to study the effects of basic factors (time, ethanol, and pH) on the extractability of polyphenolic phytochemicals from grape seeds. Total ployphenol (TP in mg gallic acid equivalents/100 g dry weight) from three types of grape seeds (Savatiano, Moschofilero, and Agiorgitiko) were recorded. The data follow.

Run	Ethanol (%)	pH	Time (H)	TP Moschofilero	Savatiano	Agiorgitiko
1	40	2	1	13,320	13,127	8,673
2	40	2	5	13,596	8,925	4,370
3	40	6	1	10,714	12,047	8,049
4	40	6	5	10,730	11,299	5,315
5	60	2	1	12,149	9,700	9,384
6	60	2	5	10,910	7,107	8,290
7	60	6	1	11,620	8,755	7,905
8	60	6	5	9,757	9,792	9,347
9	40	4	3	13,593	9,748	7,253
10	60	4	3	13,459	8,727	8,390
11	50	2	3	11,980	7,164	7,611
12	50	6	3	10,338	5,928	7,292
13	50	4	1	13,992	12,200	8,305
14	50	4	5	13,450	10,552	8,380
15	50	4	3	11,745	9,284	8,792
16	50	4	3	12,267	9,084	8,302

(a) Build a second-order model for each seed type and compare the models.
(b) Comment on the importance of any interaction terms or second-order terms in the models from part (a).
(c) Analyze the residuals from each model.

(a) Build a second order model for each seed type.

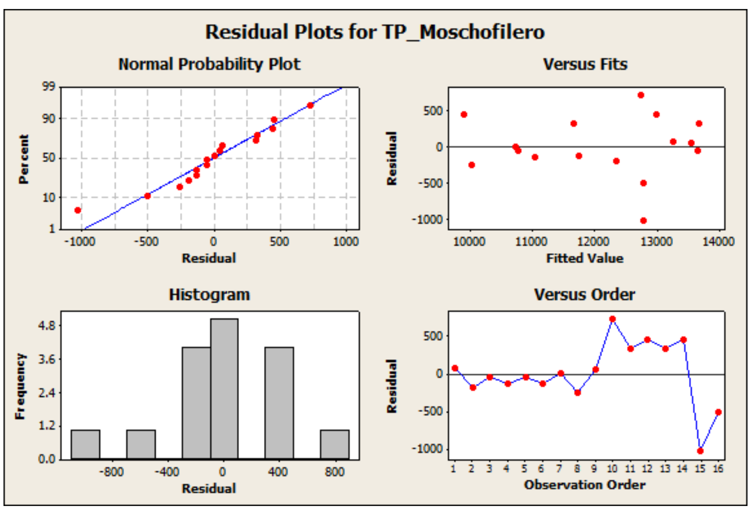
For seed type Moschofilero:

```
The analysis was done using coded units.

Estimated Regression Coefficients for TP_Moschofilero

Term                Coef   SE Coef       T      P
Constant         12776.9     312.4  40.896  0.000
Ethanol           -405.8     208.7  -1.945  0.100
pH                -879.6     208.7  -4.215  0.006
Time              -335.2     208.7  -1.606  0.159
Ethanol*Ethanol    363.7     406.4   0.895  0.405
pH*pH            -2003.3     406.4  -4.929  0.003
    Time*Time          558.7     406.4   1.375  0.218
    Ethanol*pH         473.7     233.3   2.031  0.089
    Ethanol*Time      -424.3     233.3  -1.818  0.119
    pH*Time           -110.5     233.3  -0.474  0.653


    S = 659.915    PRESS = 11008376
    R-Sq = 90.47%  R-Sq(pred) = 59.84%  R-Sq(adj) = 76.17%


    Analysis of Variance for TP_Moschofilero
    Source              DF    Seq SS    Adj SS    Adj MS      F      P
    Regression           9  24800962  24800962   2755662   6.33  0.018
      Linear             3  10507288  10507288   3502429   8.04  0.016
```

Ethanol	1	1646736	1646736	1646736	3.78	0.100
pH	1	7736962	7736962	7736962	17.77	0.006
Time	1	1123590	1123590	1123590	2.58	0.159
Square	3	10960575	10960575	3653525	8.39	0.014
Ethanol*Ethanol	1	361616	348646	348646	0.80	0.405
pH*pH	1	9776161	10580757	10580757	24.30	0.003
Time*Time	1	822797	822797	822797	1.89	0.218
Interaction	3	3333099	3333099	1111033	2.55	0.152
Ethanol*pH	1	1795512	1795512	1795512	4.12	0.089
Ethanol*Time	1	1439904	1439904	1439904	3.31	0.119
pH*Time	1	97682	97682	97682	0.22	0.653
Residual Error	6	2612927	2612927	435488		
Lack-of-Fit	5	2476685	2476685	495337	3.64	0.378
Pure Error	1	136242	136242	136242		
Total	15	27413889				

Estimated Regression Coefficients for TP_Moschofilero using data in uncoded units

Term	Coef
Constant	20627.0
Ethanol	-435.348
pH	2465.39
Time	165.542
Ethanol*Ethanol	3.63655
pH*pH	-500.836
Time*Time	139.664
Ethanol*pH	23.6875
Ethanol*Time	-21.2125
pH*Time	-27.6250



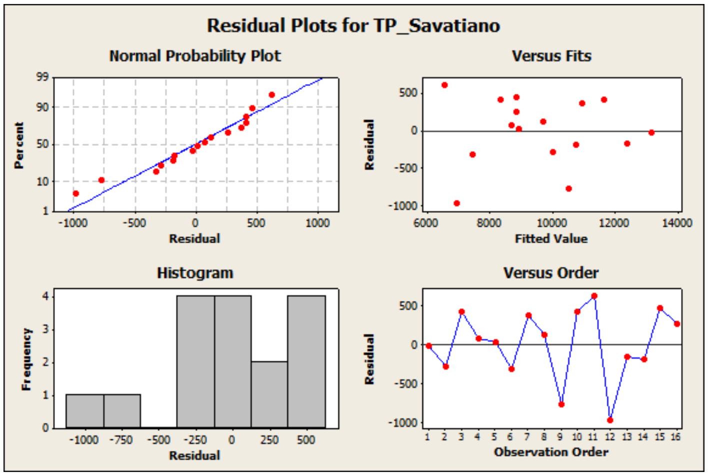
For seed type Savatiano:

The analysis was done using coded units.

Estimated Regression Coefficients for TP_Savatiano

Term	Coef	SE Coef	T	P
Constant	8824.69	333.2	26.488	0.000
Ethanol	-1106.50	222.5	-4.972	0.003
pH	179.80	222.5	0.808	0.450
Time	-815.40	222.5	-3.664	0.011

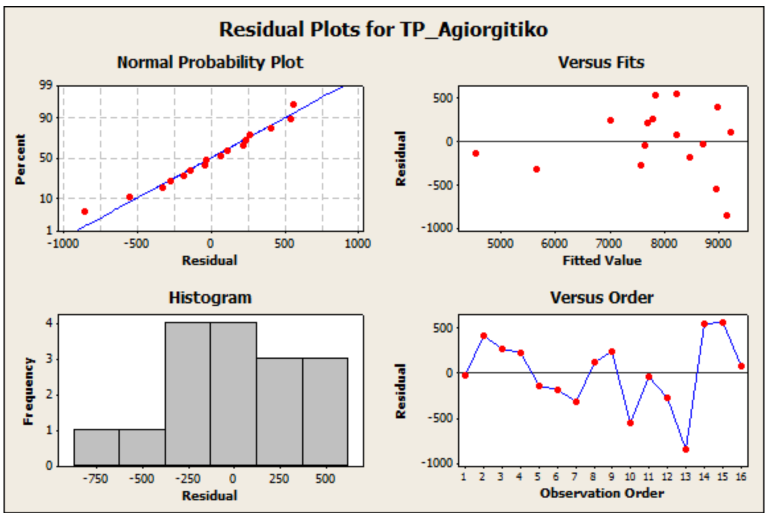
```
Ethanol*Ethanol    592.47    433.4   1.367  0.221
pH*pH            -2099.03    433.4  -4.843  0.003
Time*Time         2730.97    433.4   6.301  0.001
Ethanol*pH          55.75    248.8   0.224  0.830
Ethanol*Time       424.25    248.8   1.705  0.139
pH*Time            885.50    248.8   3.559  0.012


S = 703.718    PRESS = 22720747
R-Sq = 94.79%  R-Sq(pred) = 60.12%  R-Sq(adj) = 86.96%


Analysis of Variance for TP_Savatiano

Source               DF    Seq SS    Adj SS    Adj MS      F      P
Regression            9  54006900  54006900   6000767  12.12  0.003
  Linear              3  19215474  19215474   6405158  12.93  0.005
    Ethanol           1  12243422  12243422  12243422  24.72  0.003
    pH                1    323280    323280    323280   0.65  0.450
    Time              1   6648772   6648772   6648772  13.43  0.011
  Square              3  27053775  27053775   9017925  18.21  0.002
    Ethanol*Ethanol   1   2952824    925404    925404   1.87  0.221
    pH*pH             1   4438496  11615675  11615675  23.46  0.003
    Time*Time         1  19662455  19662455  19662455  39.70  0.001
  Interaction         3   7737651   7737651   2579217   5.21  0.042
    Ethanol*pH        1     24864     24864     24864   0.05  0.830
    Ethanol*Time      1   1439904   1439904   1439904   2.91  0.139
    pH*Time           1   6272882   6272882   6272882  12.67  0.012
Residual Error        6   2971311   2971311    495218
  Lack-of-Fit         5   2951311   2951311    590262  29.51  0.139
  Pure Error          1     20000     20000     20000
Total                15  56978211


Estimated Regression Coefficients for TP_Savatiano using data in uncoded
units

Term                  Coef
Constant           34176.7
Ethanol           -777.903
pH                 3484.47
Time              -6450.27
Ethanol*Ethanol    5.92466
pH*pH             -524.759
Time*Time          682.741
Ethanol*pH         2.78750
Ethanol*Time       21.2125
pH*Time            221.375
```



For seed type Agiorgitiko:

```
The analysis was done using coded units.
Estimated Regression Coefficients for TP_Agiorgitiko

Term                 Coef   SE Coef       T      P
Constant          8231.10     289.8  28.400  0.000
Ethanol            965.60     193.6   4.988  0.002
pH                 -42.00     193.6  -0.217  0.835
Time              -661.40     193.6  -3.417  0.014
Ethanol*Ethanol   -251.66     377.0  -0.667  0.529
pH*pH             -621.66     377.0  -1.649  0.150
Time*Time          269.34     377.0   0.714  0.502
Ethanol*pH         -92.88     216.4  -0.429  0.683
Ethanol*Time       923.13     216.4   4.265  0.005
pH*Time            513.12     216.4   2.371  0.055


S = 612.177     PRESS = 18217164
R-Sq = 91.56%   R-Sq(pred) = 31.62%   R-Sq(adj) = 78.90%


Analysis of Variance for TP_Agiorgitiko
Source                DF    Seq SS    Adj SS    Adj MS      F      P
Regression             9  24391655  24391655   2710184   7.23  0.013
  Linear               3  13715973  13715973   4571991  12.20  0.006
    Ethanol            1   9323834   9323834   9323834  24.88  0.002
    pH                 1     17640     17640     17640   0.05  0.835
    Time               1   4374500   4374500   4374500  11.67  0.014
  Square               3   1683019   1683019    561006   1.50  0.308
    Ethanol*Ethanol    1    649168    166962    166962   0.45  0.529
    pH*pH              1    842592   1018836   1018836   2.72  0.150
    Time*Time          1    191259    191259    191259   0.51  0.502
  Interaction          3   8992662   8992662   2997554   8.00  0.016
    Ethanol*pH         1     69006     69006     69006   0.18  0.683
    Ethanol*Time       1   6817278   6817278   6817278  18.19  0.005
    pH*Time            1   2106378   2106378   2106378   5.62  0.055
Residual Error         6   2248567   2248567    374761
  Lack-of-Fit          5   2128517   2128517    425703   3.55  0.382
  Pure Error           1    120050    120050    120050
Total                 15  26640222
```

```
Estimated Regression Coefficients for TP_Agiorgitiko using data in uncoded
     units

Term                  Coef
Constant           3841.29
Ethanol            228.321
pH                 1069.65
Time              -3555.65
Ethanol*Ethanol   -2.51655
pH*pH             -155.414
Time*Time          67.3362
Ethanol*pH        -4.64375
Ethanol*Time       46.1563
pH*Time            128.281
```



(b) Use $\alpha = 0.05$
For Moschofilero, the second-order term for pH is significant
For Savatiano, the second-order terms for pH and Time are important, pH*Time is important
For Agiorgitiko, Ethanol*Time and pH*Time are important

(c) The residual plots for Moschofilero and Savatiano are satisfactory. For Agiorgitiko, from the residuals versus fitted values, the variation increases with the fitted value.

14-98 Consider the data in Exercise 14-97. Use only the first eight rows of data from the 23 factorial design and assume that the experiment was conducted in blocks based on seed type.
(a) Analyze the factorial effects and comment on which effects are important.
(b) Develop a regression model to predict the response in terms of the actual factor levels.
(c) Does a residual analysis indicate any problems?

(a)
```
Estimated Effects and Coefficients for Extract (coded units)

Term               Effect     Coef  SE Coef      T      P
Constant                    9786.7    359.2  27.25  0.000
Block 1                     1812.8    507.9   3.57  0.003
Block 2                      307.3    507.9   0.60  0.555
Time              -1333.8   -666.9    359.2  -1.86  0.085
pH                 -351.8   -175.9    359.2  -0.49  0.632
Ethanol            -454.1   -227.0    359.2  -0.63  0.537
Time*pH             858.7    429.4    359.2   1.20  0.252
Time*Ethanol        615.4    307.7    359.2   0.86  0.406
```

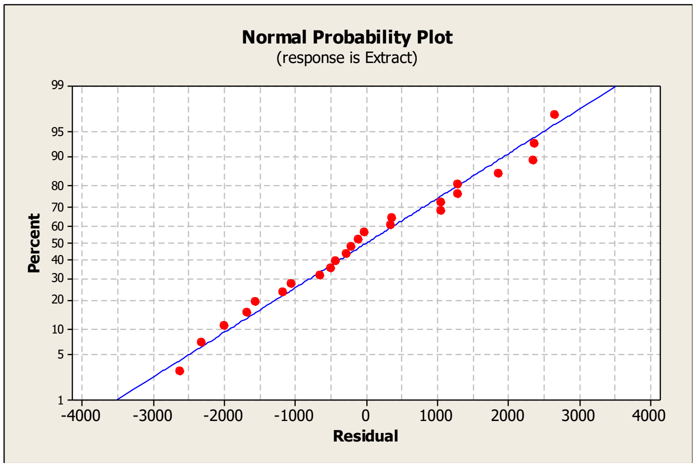
```
pH*Ethanol          291.1    145.5     359.2    0.41   0.691
Time*pH*Ethanol      64.9     32.5     359.2    0.09   0.929

S = 1759.54      PRESS = 127378161
R-Sq = 65.66%    R-Sq(pred) = 0.00%    R-Sq(adj) = 43.59%


Analysis of Variance for Extract (coded units)

Source                  DF     Seq SS    Adj SS    Adj MS       F       P
Blocks                   2   63003161  63003161  31501581   10.17   0.002
Main Effects             3   12652853  12652853   4217618    1.36   0.295
  Time                   1   10673334  10673334  10673334    3.45   0.085
  pH                     1     742368    742368    742368    0.24   0.632
  Ethanol                1    1237150   1237150   1237150    0.40   0.537
2-Way Interactions       3    7205512   7205512   2401837    0.78   0.527
  Time*pH                1    4424709   4424709   4424709    1.43   0.252
  Time*Ethanol           1    2272426   2272426   2272426    0.73   0.406
  pH*Ethanol             1     508377    508377    508377    0.16   0.691
3-Way Interactions       1      25285     25285     25285    0.01   0.929
  Time*pH*Ethanol        1      25285     25285     25285    0.01   0.929
Residual Error          14   43343958  43343958   3095997
Total                   23  126230769
```

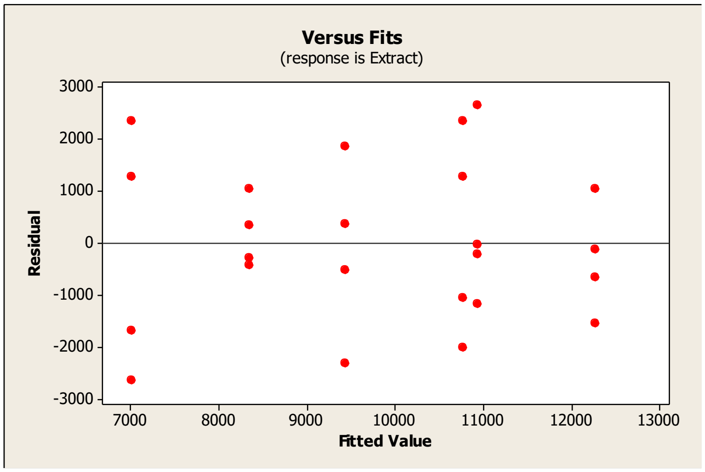
Only the main effect of Time is significant at $\alpha = 0.1$. The data used did not include the axial points used to fit a second-order model.

(b) The model based on Time is used. Blocks are important so coefficients for blocks are also shown in the following table.

```
Estimated Coefficients for Extract using data in uncoded units

Term          Coef
Constant   9786.71
Block 1    1812.79
Block 2    307.292
Time      -666.875
```

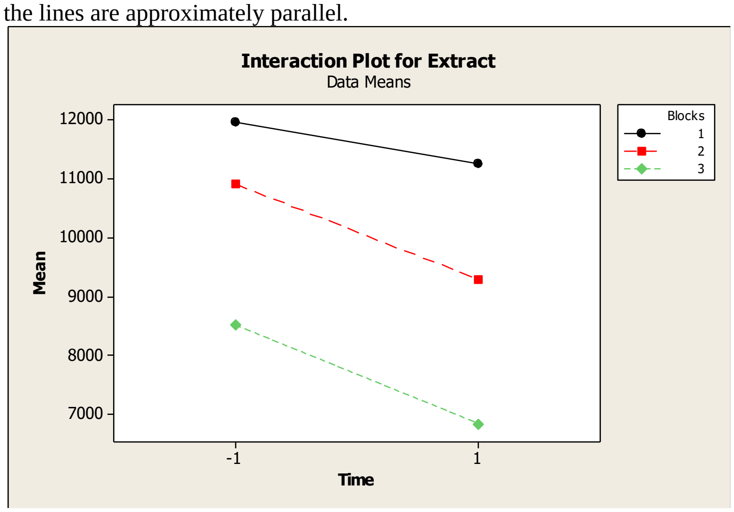
(c) Residual plots are satisfactory.



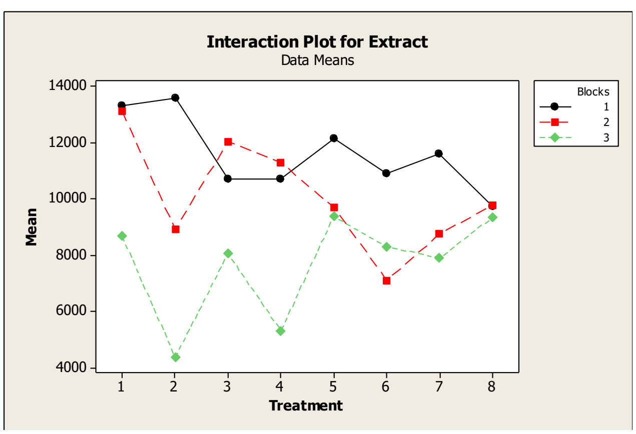


14-99 Graphically study the assumption of no interactions between blocks and treatments in the previous exercise.

The means of extract versus Time and block are plotted below. The assumption of no interaction is reasonable because the lines are approximately parallel.



However, consider the factorial experiment as 8 treatments. The means of extract for treatments and blocks are plotted below. The assumption of no interaction between treatments and blocks is a concern because the lines are not parallel.



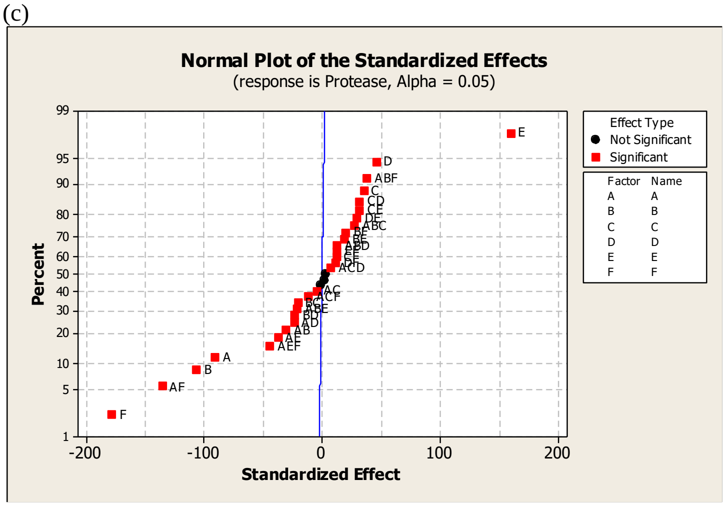
14-100 An article in *Journal of Applied Microbiology* ["Use of Response Surface Methodology to Optimize Protease Synthesis by a Novel Strain of Bacillus sp. Isolated from Portuguese Sheep Wool" (2012, Vol. 113(1), pp. 36-43)] described a fractional factorial design with three center points to study six factors (yeast extract, peptone, inoculum concentration, agitation, temperature, and pH) for protease activity. Response unitswere proteolytic activity per ml (U/ml). The data follow.

Run	Yeast (g/L)	Peptone (g/L)	Inoculum (vol.%)	Agitation (rpm)	Temperature (°C)	pH	Protease Activity (U/ml)
1	5	2	1	0	34	6	29.81
2	10	2	1	0	34	8	15.52
3	5	4	1	0	34	8	18.23
4	10	4	1	0	34	6	27.25
5	5	2	3	0	34	8	25.81
6	10	2	3	0	34	6	37.21
7	5	4	3	0	34	6	22.75
8	10	4	3	0	34	8	8.01
9	5	2	1	100	34	8	26.01
10	10	2	1	100	34	6	37.21
11	5	4	1	100	34	6	26.73
12	10	4	1	100	34	8	7.4
13	5	2	3	100	34	6	36.97
14	10	2	3	100	34	8	16.24
15	5	4	3	100	34	8	18.15
16	10	4	3	100	34	6	25.3
17	5	2	1	0	40	8	32.6
18	10	2	1	0	40	6	47.33
19	5	4	1	0	40	6	37.33
20	10	4	1	0	40	8	12.32
21	5	2	3	0	40	6	37.7
22	10	2	3	0	40	8	19.01
23	5	4	3	0	40	8	36.24
24	10	4	3	0	40	6	35.88
25	5	2	1	100	40	6	39.94
26	10	2	1	100	40	8	20.34
27	5	4	1	100	40	8	38.19
28	10	4	1	100	40	6	32.72
29	5	2	3	100	40	8	52.85
30	10	2	3	100	40	6	54.61
31	5	4	3	100	40	6	41.09
32	10	4	3	100	40	8	21.34
33	7.5	3	2	50	37	7	25.1
34	7.5	3	2	50	37	7	25.5
35	7.5	3	2	50	37	7	25.3

(a) The generator is F=ABCDE.

(b) The resolution of this design is VI.

(c)




```
Estimated Effects and Coefficients for Protease (uncoded units)

Term        Effect     Coef  SE Coef        T      P
Constant             29.315  0.03536   829.16  0.000
A           -6.419   -3.210  0.03536   -90.78  0.000
B           -7.514   -3.757  0.03536  -106.27  0.000
C            2.514    1.257  0.03536    35.56  0.001
D            3.256    1.628  0.03536    46.04  0.000
E           11.306    5.653  0.03536   159.89  0.000
F          -12.598   -6.299  0.03536  -178.16  0.000
A*B         -2.142   -1.071  0.03536   -30.29  0.001
A*C         -0.326   -0.163  0.03536    -4.61  0.044
A*D         -1.677   -0.838  0.03536   -23.71  0.002
A*E         -2.629   -1.315  0.03536   -37.18  0.001
A*F         -9.568   -4.784  0.03536  -135.31  0.000
B*C         -1.441   -0.720  0.03536   -20.37  0.002
B*D         -1.642   -0.821  0.03536   -23.22  0.002
B*E          1.356    0.678  0.03536    19.17  0.003
B*F          1.452    0.726  0.03536    20.53  0.002
C*D          2.237    1.118  0.03536    31.63  0.001
C*E          2.229    1.115  0.03536    31.53  0.001
C*F          0.866    0.433  0.03536    12.24  0.007
D*E          2.078    1.039  0.03536    29.39  0.001
D*F          0.842    0.421  0.03536    11.91  0.007
E*F          0.884    0.442  0.03536    12.51  0.006
A*B*C        1.962    0.981  0.03536    27.75  0.001
A*B*D        0.888    0.444  0.03536    12.56  0.006
A*B*E       -1.457   -0.728  0.03536   -20.60  0.002
A*B*F        2.694    1.347  0.03536    38.10  0.001
A*C*D        0.529    0.265  0.03536     7.49  0.017
A*C*E        0.114    0.057  0.03536     1.62  0.247
A*C*F       -0.799   -0.400  0.03536   -11.30  0.008
A*D*E       -0.039   -0.020  0.03536    -0.56  0.634
A*D*F        0.194    0.097  0.03536     2.75  0.111
A*E*F       -3.101   -1.550  0.03536   -43.85  0.001
Ct Pt                -4.015  0.12076   -33.25  0.001

S = 0.2          PRESS = *
R-Sq = 100.00%   R-Sq(pred) = *%   R-Sq(adj) = 99.97%


Analysis of Variance for Protease (uncoded units)
```

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	6	3209.00	3209.00	534.83	13370.84	0.000
A	1	329.67	329.67	329.67	8241.68	0.000
B	1	451.73	451.73	451.73	11293.17	0.000
C	1	50.58	50.58	50.58	1264.42	0.001
D	1	84.79	84.79	84.79	2119.82	0.000
E	1	1022.54	1022.54	1022.54	25563.43	0.000
F	1	1269.70	1269.70	1269.70	31742.55	0.000
2-Way Interactions	15	1049.74	1049.74	69.98	1749.57	0.001
A*B	1	36.70	36.70	36.70	917.53	0.001
A*C	1	0.85	0.85	0.85	21.21	0.044
A*D	1	22.50	22.50	22.50	562.38	0.002
A*E	1	55.31	55.31	55.31	1382.72	0.001
A*F	1	732.39	732.39	732.39	18309.80	0.000
B*C	1	16.60	16.60	16.60	415.08	0.002
B*D	1	21.57	21.57	21.57	539.15	0.002
B*E	1	14.70	14.70	14.70	367.54	0.003
B*F	1	16.86	16.86	16.86	421.59	0.002
C*D	1	40.03	40.03	40.03	1000.72	0.001
C*E	1	39.76	39.76	39.76	994.02	0.001
C*F	1	5.99	5.99	5.99	149.86	0.007
D*E	1	34.55	34.55	34.55	863.72	0.001
D*F	1	5.67	5.67	5.67	141.75	0.007
E*F	1	6.26	6.26	6.26	156.42	0.006
3-Way Interactions	10	196.84	196.84	19.68	492.11	0.002
A*B*C	1	30.79	30.79	30.79	769.79	0.001
A*B*D	1	6.31	6.31	6.31	157.75	0.006
A*B*E	1	16.98	16.98	16.98	424.50	0.002
A*B*F	1	58.08	58.08	58.08	1451.93	0.001
A*C*D	1	2.24	2.24	2.24	56.05	0.017
A*C*E	1	0.10	0.10	0.10	2.62	0.247
A*C*F	1	5.11	5.11	5.11	127.80	0.008
A*D*E	1	0.01	0.01	0.01	0.31	0.634
A*D*F	1	0.30	0.30	0.30	7.56	0.111
A*E*F	1	76.91	76.91	76.91	1922.78	0.001
Curvature	1	44.22	44.22	44.22	1105.56	0.001
Residual Error	2	0.08	0.08	0.04		
Pure Error	2	0.08	0.08	0.04		
Total	34	4499.89				

Although many effects are significant based on the error term from the center points, the major effects from the normal probability plot are A, B, E, F, and AF. A reduced model for these effects is constructed.

Estimated Effects and Coefficients for Protease (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		28.971	0.8268	35.04	0.000
Yeast	-6.419	-3.210	0.8647	-3.71	0.001
Peptone	-7.514	-3.757	0.8647	-4.35	0.000
Temperature	11.306	5.653	0.8647	6.54	0.000
pH	-12.598	-6.299	0.8647	-7.28	0.000
Yeast*pH	-9.568	-4.784	0.8647	-5.53	0.000

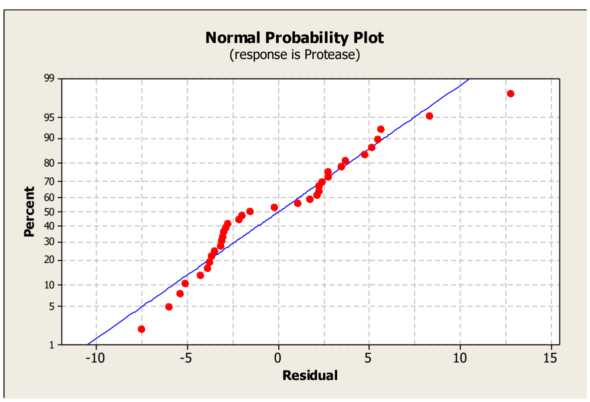
S = 4.89145 PRESS = 1026.13
R-Sq = 84.58% R-Sq(pred) = 77.20% R-Sq(adj) = 81.92%

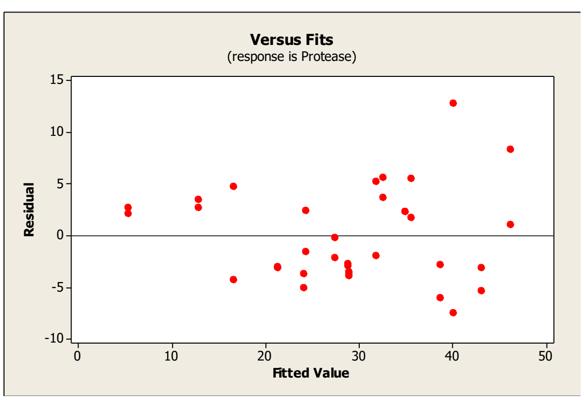
Analysis of Variance for Protease (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	3073.63	3073.63	768.41	32.12	0.000
Yeast	1	329.67	329.67	329.67	13.78	0.001

```
  Peptone              1   451.73   451.73   451.73  18.88  0.000
  Temperature          1  1022.54  1022.54  1022.54  42.74  0.000
  pH                   1  1269.70  1269.70  1269.70  53.07  0.000
2-Way Interactions     1   732.39   732.39   732.39  30.61  0.000
  Yeast*pH             1   732.39   732.39   732.39  30.61  0.000
Residual Error        29   693.86   693.86    23.93
  Curvature            1    44.22    44.22    44.22   1.91  0.178
  Lack of Fit         10   324.07   324.07    32.41   1.79  0.135
  Pure Error          18   325.57   325.57    18.09
Total                 34  4499.89

Unusual Observations for Protease

Obs  StdOrder  Protease      Fit  SE Fit  Residual  St Resid
 29        29   52.8500  40.0758  2.1029   12.7742     2.89R

R denotes an observation with a large standardized residual.

Estimated Coefficients for Protease using data in uncoded units

Term              Coef
Constant      -76.2181
Yeast          12.1115
Peptone       -3.75719
Temperature    1.88427
pH             8.05313
Yeast*pH      -1.91363
```

The residual plots from the reduced model follow. The plot of residuals versus fitted values exhibits some non-constant variance or curvature. A transformation might be considered.





<u>Mind Expanding Exercises</u>

14-101 Consider an unreplicated 2^k factorial, and suppose that one of the treatment combinations is missing. One logical approach to this problem is to estimate the missing value with a number that makes the highest order interaction estimate zero. Apply this technique to the data in Example 14-5, assuming that *ab* is missing. Compare the results of the analysis of the application of these data with the results in Example 14-5.

The ABCD interaction is

$$\frac{1}{8}\left[(1)+ab+ac+bc+ad+bd+cd+abcd\right]-\left[a+b+c+d+abc+abd+acd+bcd\right]$$

If *ab* is missing, then ABCD interaction will be zero when
[550 + ab + 642 + 601 + 749 + 1052 + 1075 + 729] – [669 + 604 + 633 + 1037 + 635 + 868 + 860 + 1063] = 0
Therefore, *ab* + 5398 - 6369 = 0 or *ab* = 971. After estimating *ab*, only the A and AD effects appear significant.

14-102 What blocking scheme would you recommend if it were necessary to run a 2^4 design in four blocks of four runs each?

Two three-factor interactions could be used to generate the blocks such as ABC and ACD. This would confound these effects and ABC(ACD) = BD with blocks. Therefore, only one two-factor and no main effects are confounded with blocks.

14-103 Consider a 2^2 design in two blocks with *AB* confounded with blocks. Prove algebraically that $SS_{\text{AB}} = SS_{\text{Blocks}}$.

	A	**B**	**AB**	**block**
(1)	-	-	+	1
a	+	-	-	2
b	-	+	-	2
ab	+	+	+	1

The block effect is estimated by $\frac{a+b}{2}-\frac{(1)-ab}{2}$ which is the same as the estimate of the effect of AB.

14-104 Consider a 2^3 design. Suppose that the largest number of runs that can be made in one block is four, but you can afford to perform a total of 32 observations.

(a) Suggest a blocking scheme will provides some information on all interactions.
(b) Show an outline (source of variability, degrees of freedom only) for the analysis of variance for this design.

(a) A different effect can be confounded in each replicate as follows.

Replicate 1		Replicate 2		Replicate 3		Replicate 4	
ABC confounded		AB confounded		BC confounded		AC confounded	
(1)	a	(1)	a	(1)	b	(1)	a
ab	b	c	b	a	c	b	c
ac	c	ab	ac	bc	ab	ac	ab
bc	abc	abc	bc	abc	ac	abc	bc

(b)

Source of Variation	Degrees of freedom
Replicates	3
Blocks with replicates	4
[or ABC (rep. 1) + AB (rep. 2) + BC (rep. 3) + AC (rep. 4)]	
A	1
B	1
C	1
AB (from replicates 1, 3, and 4)	1
AC (from replicates 1, 2, and 3)	1
BC (from replicates 1, 2, and 4)	1
ABC (from replicates 2, 3, and 4)	1
Error (by subtraction)	17
Total	31

In calculating an interaction sum of squares, only data from the replicates in which the interaction is un-confounded are used.

14-105 Construct a 2^{5-1} design. Suppose that it is necessary to run this design in two blocks of eight runs each. Show how this can be done by confounding a two-factor interaction (and its aliased three-factor interaction) with blocks.

A	B	C	D	E = ABCD	AB = CDE	block
-	-	-	-	+	+	1
+	-	-	-	-	-	2
-	+	-	-	-	-	2
+	+	-	-	+	+	1
-	-	+	-	-	+	1
+	-	+	-	+	-	2
-	+	+	-	+	-	2
+	+	+	-	-	+	1
-	-	-	+	-	+	1
+	-	-	+	+	-	2
-	+	-	+	+	-	2
+	+	-	+	-	+	1
-	-	+	+	+	+	1
+	-	+	+	-	-	2
-	+	+	+	-	-	2
+	+	+	+	+	+	1

This uses AB = CDE as the effect to confound with blocks.

14-106 Construct a 2_{IV}^{7-2} design. Show how this design may be confounded in four blocks of eight runs each. Are any two-factor interactions confounded with blocks?

The generators are F = ABCD and G = ABDE. The complete defining relation is
I = ABCDF = ABDEG = CEFG.
The design can be constructed in four blocks by confounding ACE = AFG and BCE = BFG with blocks. This also confounds AB = CDF = DEG with blocks.
Yes, a two-factor interaction is confounded with blocks. The best blocking scheme confounds only one two-factor interaction with blocks.

14-107 Construct a 2_{IV}^{7-3} design. Show how this design can be confounded in two blocks of eight runs each without losing information on any of the two-factor interactions.

The generators are E = ABC, F = BCD, and G = ACD.
The complete defining relation is
I = ABCE = BCDF = ADEF = ACDG = BDEG = ABFG = CEFG.
The alias set
ABD = CDE = ACF = BEF = BCG = AEG = DFG
can be used to construct the blocks. Then, only three-factor interactions are confounded with blocks.

14-108 Set up a 2_{III}^{7-4} design using *D* = *AB*, *E* = *AC*, *F* = *BC*, and *G* = *ABC* as the design generators. Ignore all interactions above two factors.

(a) Verify that each main effect is aliased with three two factor interactions.
(b) Suppose that a second 2_{III}^{7-4} design with generators *D* = –*AB*, *E* = –*AC*, *F* = –*BC*, and *G* = *ABC* is run. What are the aliases of the main effects in this design?
(c) What factors may be estimated if the two sets of factor effect estimates above are combined?

(a)

A	B	C	D = AB	E = AC	F = BC	G = ABC
-	-	-	+	+	+	-
+	-	-	-	-	+	+
-	+	-	-	+	-	+
+	+	-	+	-	-	-
-	-	+	+	-	-	+
+	-	+	-	+	-	-
-	+	+	-	-	+	-
+	+	+	+	+	+	+

The complete defining relation is
I = ABD = ACE = BCDE = BCF = ACDF = ABEF = DEF = ABCG = CDG = BEG = ADEG = AFG = BDFG = CEFG = ABCDEFG.

The alias structure follows (including only one- and two-factor effects).
A = BD = CE = FG
B = AD = CF = EG
C = AE = BF = DG
D = AB = EF = CG
E = AC = DF = BG
F = BC = DE = AG
G = CD = BE = AF

(b) The complete defining relation is
I = -ABD = -ACE = BCDE = -BCF = ACDF = ABEF = -DEF = ABCG = -CDG = -BEG = ADEG = -AFG = BDFG = CEFG = -ABCDEFG.

The aliases (up to two-factor effects) are:
A = -BD = -CE = -FG
B = -AD = -CF = -EG
C = -AE = -BF = -DG
D = -AB = -EF = -CG
E = -AC = -DF = -BG
F = -BC = -DE = -AG
G = -CD = -BE = -AF

(c) All the main effects can be estimated. Use the average response from the alias set that contains the main effect in each fraction. The two-factor effects cancel when this average is computed.

14-109 Consider the square root of the sum of squares for curvature and divide by the square root of mean square error. Explain why the statistic that results has a *t* distribution and why it can be used to conduct a *t* test for curvature that is equivalent to the *F* test in the ANOVA.

When the square root of the sum of squares for curvature is divided by the square root of mean squared error, the resulting statistic is

$$\frac{|\,\bar{y}_F - \bar{y}_C\,|}{\hat{\sigma}\sqrt{\dfrac{1}{n_F} + \dfrac{1}{n_C}}}$$

and this is a t-statistic used to compare two means. If this t-statistic is significant, $\bar{y}_F - \bar{y}_C$ is large meaning curvature is significant. This test is equivalent to the F test from the ANOVA testing for curvature. This statistic is compared to a t distribution with the degrees of freedom associated with the estimate of σ.

If a random variable with a *t* distribution is squared, the resulting random variable has an *F* distribution with one degree of freedom in the numerator and degrees of freedom in the denominator equal to the degrees of freedom of the *t* statistic. Therefore, the reference distribution for the *F* test is the square of the reference distribution for the *t* test.

CHAPTER 15

Section 15-3

15-1 Control charts for $\bar{X}$ and R are to be set up for an important quality characteristic. The sample size is n = 5, and $\bar{x}$ and r are computed for each of 35 preliminary samples. The summary data are

$$\sum_{i=1}^{35} \bar{x}_i = 7805 \qquad \sum_{i=1}^{35} r_i = 1200$$

(a) Calculate trial control limits for $\bar{X}$ and R charts.
(b) Assuming that the process is in control, estimate the process mean and standard deviation.

(a) $\bar{\bar{x}} = \dfrac{7805}{35} = 223 \qquad \bar{r} = \dfrac{1200}{35} = 34.286$

$\bar{x}$ chart

$$UCL = CL + A_2\bar{r} = 223 + 0.577(34.286) = 242.78$$
$$CL = 223$$
$$LCL = CL - A_2\bar{r} = 223 - 0.577(34.286) = 203.22$$

R chart

$$UCL = D_4\bar{r} = 2.115(34.286) = 72.51$$
$$CL = 34.286$$
$$LCL = D_3\bar{r} = 0(34.286) = 0$$

(b)

$$\hat{\mu} = \bar{\bar{x}} = 223$$
$$\hat{\sigma} = \frac{\bar{r}}{d_2} = \frac{34.286}{2.326} = 14.74$$

15-2 Twenty-five samples of size 5 are drawn from a process at one-hour intervals, and the following data are obtained:

$$\sum_{i=1}^{25} \bar{x}_i = 362.75 \qquad \sum_{i=1}^{25} r_i = 8.60 \qquad \sum_{i=1}^{25} s_i = 3.64$$

(a) Calculate trial control limits for $\bar{X}$ and R charts.
(b) Repeat part (a) for $\bar{X}$ and S charts.

(a) $\bar{\bar{x}} = \dfrac{362.75}{25} = 14.510 \qquad \bar{r} = \dfrac{8.60}{25} = 0.344 \quad \bar{s} = \dfrac{3.64}{25} = 0.1456$

$\bar{x}$ *chart*

$$UCL = CL + A_2\bar{r} = 14.510 + 0.577(0.344) = 14.708$$
$$CL = 14.510$$
$$LCL = CL - A_2\bar{r} = 14.510 - 0.577(0.344) = 14.312$$

R *chart*

$$UCL = D_4\bar{r} = 2.115(0.344) = 0.728$$
$$CL = 0.344$$
$$LCL = D_3\bar{r} = 0(0.344) = 0$$

(b) c_4=0.94

$\bar{x}$ *chart*

$$UCL = CL + 3\frac{\bar{s}}{c_4\sqrt{n}} = 14.510 + 3\frac{0.1456}{0.94\sqrt{25}} = 14.603$$
$$CL = 14.510$$
$$LCL = CL - 3\frac{\bar{s}}{c_4\sqrt{n}} = 14.510 - 3\frac{0.1456}{0.94\sqrt{25}} = 14.417$$

S *chart*

$$UCL = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 0.1456 + 3\left(\frac{0.1456}{0.94}\right)\sqrt{1-0.94^2} = 0.3041$$
$$CL = 0.1456$$
$$LCL = \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 0.1456 - 3\left(\frac{0.1456}{0.94}\right)\sqrt{1-0.94^2} = -0.0129 \to 0$$

15-3 Control charts are to be constructed for samples of size n = 4, and x and s are computed for each of 20 preliminary samples as follows:

$$\sum_{i=1}^{20}\bar{x}_i = 4460 \qquad \sum_{i=1}^{20} s_i = 271.6$$

(a) Calculate trial control limits for $\bar{X}$ and S charts.
(b) Assuming the process is in control, estimate the process mean and standard deviation.

(a) $\bar{\bar{x}} = \dfrac{4460}{20} = 223 \qquad \bar{s} = \dfrac{271.6}{20} = 13.58$

$\bar{x}$ *chart*

$$UCL = CL + 3\frac{\bar{s}}{c_4\sqrt{n}} = 223 + 3\left(\frac{13.58}{0.9213\sqrt{4}}\right) = 245.11$$

$$CL = 223$$

$$LCL = CL - 3\frac{\bar{s}}{c_4\sqrt{n}} = 223 - 3\left(\frac{13.58}{0.9213\sqrt{20}}\right) = 200.89$$

S *chart*

$$UCL = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 13.58 + 3\left(\frac{13.58}{0.9213}\right)\sqrt{0.1512} = 30.77$$

$$CL = 13.58$$

$$LCL = \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 13.58 - 3\left(\frac{13.58}{0.9213}\right)\sqrt{0.1512} = -3.61 \rightarrow 0$$

(b) Process mean and standard deviation

$$\hat{\mu} = \bar{\bar{x}} = 223 \qquad \hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{13.58}{0.9213} = 14.74$$

15-4 Samples of size $n = 6$ are collected from a process every hour. After 20 samples have been collected, we calculate $\bar{\bar{x}} = 20.0$ and $\bar{r}/d_2 = 1.4$.

(a) Calculate trial control limits for $\bar{X}$ and R charts.

(b) If $\bar{s}/c_4 = 1.5$, calculate trial control limits for $\bar{X}$ and S charts.

(a) $\bar{\bar{x}} = 20.0 \qquad \frac{\bar{r}}{d_2} = 1.4 \qquad d_2 = 2.534 \qquad \bar{r} = 1.4(2.534) = 3.5476$

$\bar{x}$ *chart*

$$UCL = CL + A_2\bar{r} = 20.0 + 0.483(3.5476) = 21.71$$

$$CL = 20.0$$

$$LCL = CL - A_2\bar{r} = 20.0 - 0.483(3.5476) = 18.29$$

R *chart*

$$UCL = D_4\bar{r} = 2.004(3.53476) = 7.11$$

$$CL = 3.5476$$

$$LCL = D_3\bar{r} = 0(3.5476) = 0$$

$\bar{x}$ *chart* where $\bar{s}/c_4 = 1.5$

$$UCL = CL + 3\frac{\bar{s}}{c_4\sqrt{n}} = 20.0 + 3(1.5)/\sqrt{6} = 21.84$$

$$CL = 20.0$$

$$LCL = CL - 3\frac{\bar{s}}{c_4\sqrt{n}} = 20.0 - 3(1.5)/\sqrt{6} = 18.16$$

S *chart* $c_4 = 0.9515$ so $\bar{s} = 1.427$

$$UCL = CL + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 1.427 + 3(1.5)\sqrt{0.0946} = 2.811$$

$$CL = 1.427$$

$$LCL = CL - 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 1.427 - 3(1.5)\sqrt{0.0946} = 0.043$$

15-5 The level of cholesterol (in mg/dL) is an important index for human health. The sample size is n = 5. The following summary statistics are obtained from cholesterol measurements:

$$\sum_{i=1}^{30}\bar{x}_i = 140.03 \quad \sum_{i=1}^{30} r_i = 13.63 \quad \sum_{i=1}^{30} s_i = 5.10$$

(a) Find trial control limits for $\bar{X}$ and R charts.
(b) Repeat part (a) for $\bar{X}$ and S charts.

(a) $\bar{\bar{x}} = \frac{140.03}{30} = 4.6677$ $\bar{r} = \frac{13.63}{30} = 0.4543$

For the $\bar{x}$ chart:

$$UCL = \bar{\bar{x}} + A_2\bar{r} = 4.6677 + (0.577)(0.4543) = 4.930$$
$$LCL = \bar{\bar{x}} - A_2\bar{r} = 4.6677 - (0.577)(0.4543) = 4.406$$

For the *r* chart:

$$UCL = D_4\bar{r} = (2.115)(0.4543) = 0.961$$
$$LCL = D_3\bar{r} = (0)(0.4543) = 0$$

(b) $\bar{s} = \frac{5.10}{30} = 0.17$

For the $\bar{x}$ chart:

$$UCL = \bar{\bar{x}} + \frac{3\bar{s}}{c_4\sqrt{n}} = 4.6677 + \frac{3(0.17)}{0.94\sqrt{5}} = 4.910$$
$$LCL = \bar{\bar{x}} - \frac{3\bar{s}}{c_4\sqrt{n}} = 4.6677 - \frac{3(0.17)}{0.94\sqrt{5}} = 4.425$$

For the *s* chart:

$$UCL = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 0.17 + 3\left(\frac{0.17}{0.94}\right)\sqrt{1-0.94^2} = 0.355$$
$$LCL = \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} = 0.17 - 3\left(\frac{0.17}{0.94}\right)\sqrt{1-0.94^2} = -0.015$$

Because the LCL is negative it is set to zero.

15-6 An $\overline{X}$ control chart with three-sigma control limits has UCL = 48.75 and LCL = 42.71. Suppose that the process standard deviation is σ = 2.25. What subgroup size was used for the chart?

For the $\bar{x}$ chart:

$$UCL = \mu + \frac{3\sigma}{\sqrt{n}} = 48.75$$

$$LCL = \mu - \frac{3\sigma}{\sqrt{n}} = 42.71$$

The difference $UCL - LCL = \frac{6\sigma}{\sqrt{n}} = 48.75 - 42.71 = 6.04$

$$n = \left(\frac{6(2.25)}{6.04}\right)^2 = 5$$

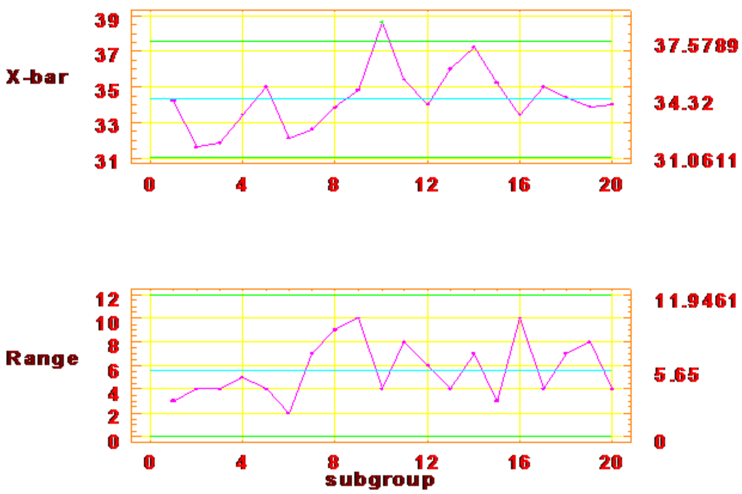
15-7 An extrusion die is used to produce aluminum rods. The diameter of the rods is a critical quality characteristic. The following table shows $\bar{x}$ and r values for 20 samples of five rods each. Specifications on the rods are 0.5035 ± 0.0010 inch. The values given are the last three digits of the measurement; that is, 34.2 is read as 0.50342.

Sample	$\bar{x}$	r
1	34.2	3
2	31.6	4
3	31.8	4
4	33.4	5
5	35.0	4
6	32.1	2
7	32.6	7
8	33.8	9
9	34.8	10
10	38.6	4
11	35.4	8
12	34.0	6
13	36.0	4
14	37.2	7
15	35.2	3
16	33.4	10
17	35.0	4
18	34.4	7
19	33.9	8
20	34.0	4

(a) Using all the data, find trial control limits for $\overline{X}$ and R charts, construct the chart, and plot the data.
(b) Use the trial control limits from part (a) to identify out-of- control points. If necessary, revise your control limits, assuming that any samples that plot outside the control limits can be eliminated. Estimate σ.

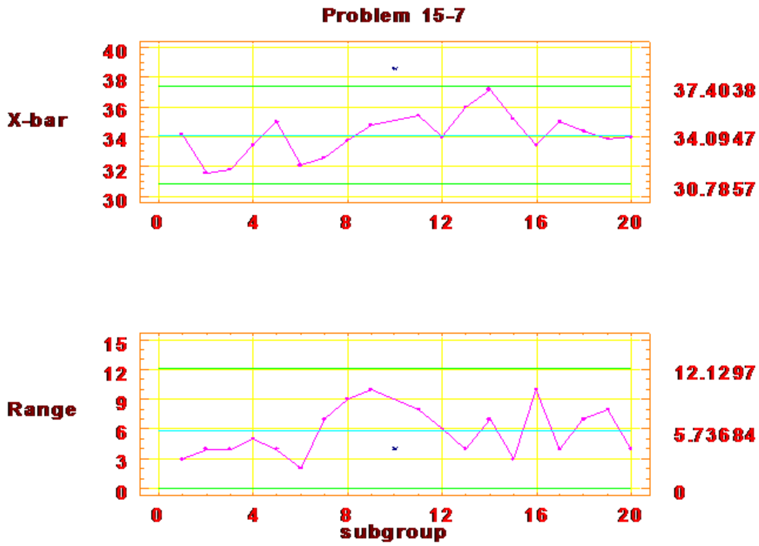
(a)
```
X-bar and Range - Initial Study
Problem 15-7
```

```
X-bar                                       | Range
-----                                       | -----
UCL:  +     3.0 sigma = 37.5789             | UCL:  +     3.0 sigma = 11.9461
Centerline            = 34.32               | Centerline            =  5.65
LCL:  -     3.0 sigma = 31.0611             | LCL:  -     3.0 sigma =  0
|
out of limits = 1                           | out of limits = 0
--------------------------------------------------------------------------------
Chart: Both      Normalize: No

   20 subgroups, size 5                         0 subgroups excluded

Estimated
process mean  = 34.32
process sigma =  2.42906
mean Range    =  5.65
```



(b)

```
                                     Charting xbar
X-bar                                       | Range
-----                                       | -----
UCL:  +     3.0 sigma = 37.4038             | UCL:  +     3.0 sigma = 12.1297
Centerline            = 34.0947             | Centerline            =  5.73684
LCL:  -     3.0 sigma = 30.7857             | LCL:  -     3.0 sigma =  0
                                            |
out of limits = 0                           | out of limits = 0
--------------------------------------------------------------------------------
20 subgroups, size 5                     1 subgroups excluded
Estimated
process mean  = 34.0947
process sigma =  2.4664
mean Range    =  5.73684
```



15-8 The copper content of a plating bath is measured three times per day, and the results are reported in ppm. The $\bar{x}$ and r values for 25 days are shown in the following table:

Day	$\bar{x}$	r	Day	$\bar{x}$	r
1	5.45	1.21	14	7.01	1.45
2	5.39	0.95	15	5.83	1.37
3	6.85	1.43	16	6.35	1.04
4	6.74	1.29	17	6.05	0.83
5	5.83	1.35	18	7.11	1.35
6	7.22	0.88	19	7.32	1.09
7	6.39	0.92	20	5.90	1.22
8	6.50	1.13	21	5.50	0.98
9	7.15	1.25	22	6.32	1.21
10	5.92	1.05	23	6.55	0.76
11	6.45	0.98	24	5.90	1.20
12	5.38	1.36	25	5.95	1.19
13	6.03	0.83			

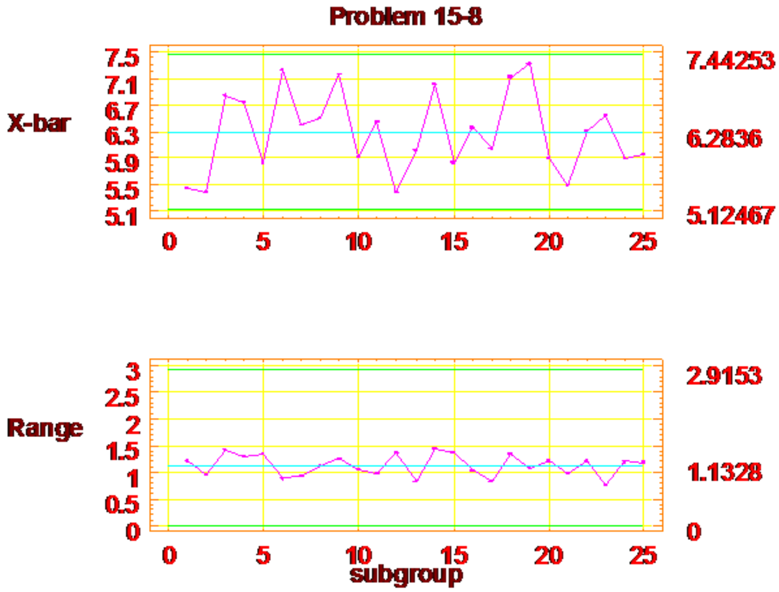
(a) Using all the data, find trial control limits for $\bar{X}$ and R charts, construct the chart, and plot the data. Is the process in statistical control?

(b) If necessary, revise the control limits computed in part (a), assuming that any samples that plot outside the control limits can be eliminated.

(a)

```
                          X-bar and Range - Initial Study
Charting Problem 15-8
X-bar                                    | Range
-----                                    | -----
UCL:  +     3.0 sigma = 7.44253          | UCL:  +     3.0 sigma = 2.9153
Centerline            = 6.2836           | Centerline            = 1.1328
LCL:  -     3.0 sigma = 5.12467          | LCL:  -     3.0 sigma = 0
                                         |
out of limits = 0                        | out of limits = 0
Estimated
process mean  = 6.2836
```

```
process sigma = 0.669108
mean Range    = 1.1328
```



There are no points beyond the control limits. The process appears to be in control.

(b) No points fell beyond the control limits. The limits do not need to be revised.

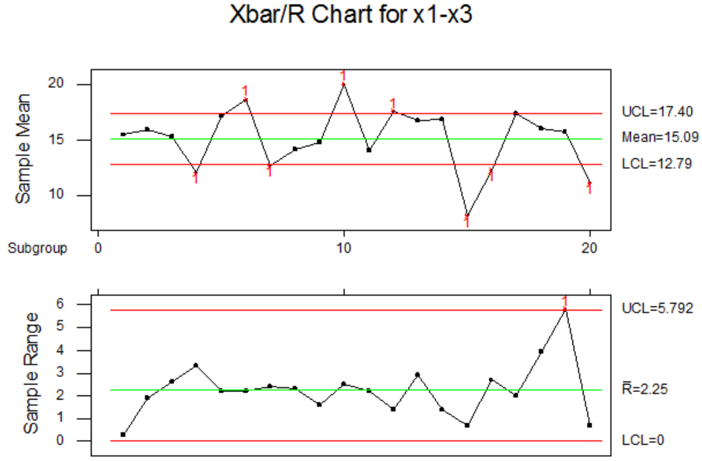
15-9 The pull strength of a wire-bonded lead for an integrated circuit is monitored. The following table provides data for 20 samples each of size 3.

Sample Number	x_1	x_2	x_3
1	15.4	15.6	15.3
2	15.4	17.1	15.2
3	16.1	16.1	13.5
4	13.5	12.5	10.2
5	18.3	16.1	17.0
6	19.2	17.2	19.4
7	14.1	12.4	11.7
8	15.6	13.3	13.6
9	13.9	14.9	15.5
10	18.7	21.2	20.1
11	15.3	13.1	13.7
12	16.6	18.0	18.0
13	17.0	15.2	18.1
14	16.3	16.5	17.7
15	8.4	7.7	8.4
16	11.1	13.8	11.9
17	16.5	17.1	18.5
18	18.0	14.1	15.9
19	17.8	17.3	12.0
20	11.5	10.8	11.2

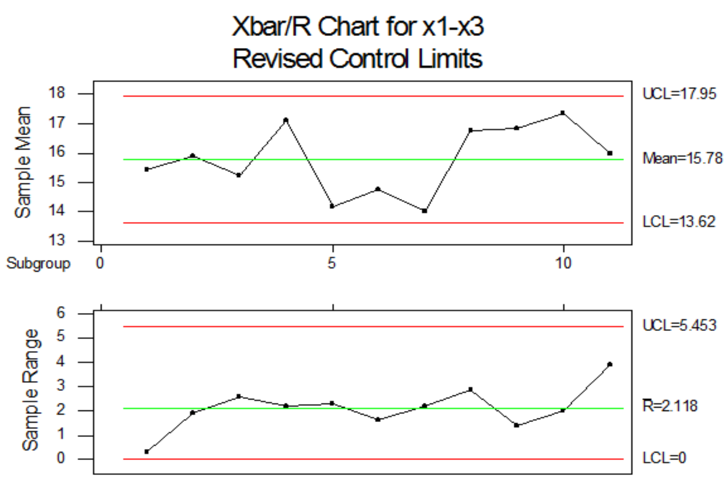
(a) Use all the data to determine trial control limits for $\bar{X}$ and R charts, construct the control limits, and plot the data.
(b) Use the control limits from part (a) to identify out-of-control points. If necessary, revise your control limits assuming that any samples that plot outside of the control limits can be eliminated.
(c) Repeat parts (a) and (b) for $\bar{X}$ and S charts.

(a)

```
                        X-bar and Range - Initial Study
            Charting Problem 15-9
X-bar                                 | Range
-----                                 | -----
UCL:  +    3.0 sigma = 17.4           | UCL:  +    3.0 sigma =  5.792
Centerline           = 15.09          | Centerline           =  2.25
LCL:  -    3.0 sigma = 12.79          | LCL:  -    3.0 sigma =  0
                                      |

Test Results: X-bar One point more than 3.00 sigmas from center line.
Test Failed at points: 4 6 7 10 12 15 16 20

Test Results for R Chart: One point more than 3.00 sigmas from center line.
Test Failed at points: 19
```



(b) Removed points 4, 6, 7, 10, 12, 15, 16, 19, and 20 and revised the control limits. The control limits are not as wide after being revised: X-bar UCL=17.96, CL=15.78, LCL=13.62 and R UCL = 5.453, R-bar = 2.118, LCL=0.

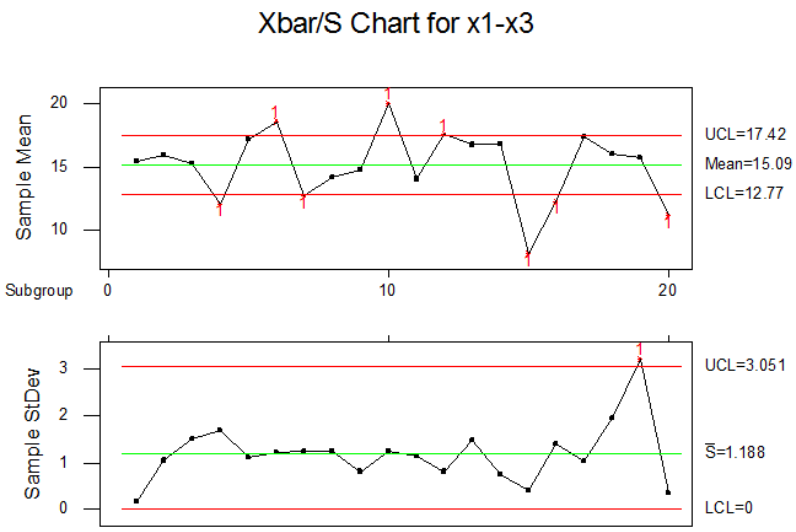


(c)

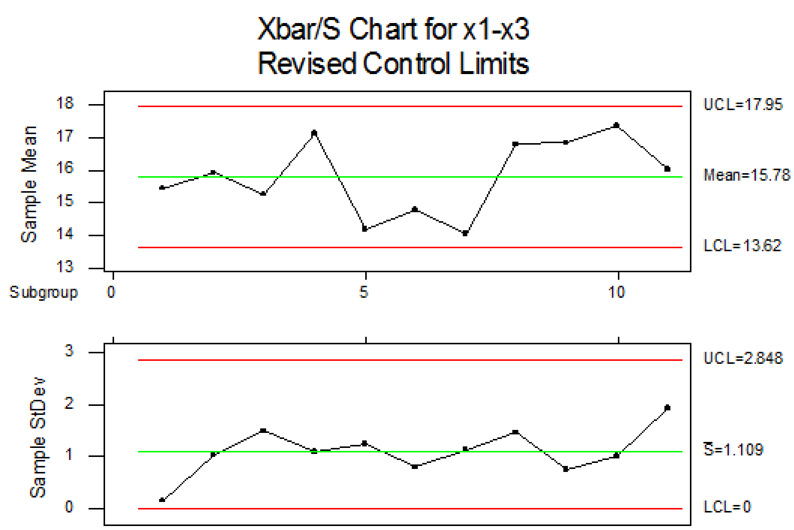
```
                        X-bar and StDev - Initial Study
                          Charting Problem 16-7
```

```
X-bar                                     | StDev
-----                                     | -----
UCL:  +    3.0 sigma = 17.42              | UCL:  +    3.0 sigma =  3.051
Centerline           = 15.09              | Centerline           =  1.188
LCL:  -    3.0 sigma = 12.77              | LCL:  -    3.0 sigma =  0
                                          |

Test Results: X-bar One point more than 3.00 sigmas from center line.
Test Failed at points: 4 6 7 10 12 15 16 20

Test Results for S Chart:One point more than 3.00 sigmas from center line.
Test Failed at points: 19
```

.



Removed points 4, 6, 7, 10, 12, 15, 16, 19, and 20 and revised the control limits. The control limits are not as wide after being revised: X-bar UCL = 17.95, CL =1 5.78, LCL=13.62 and S UCL = 2.848, S-bar=1.109, LCL=0.

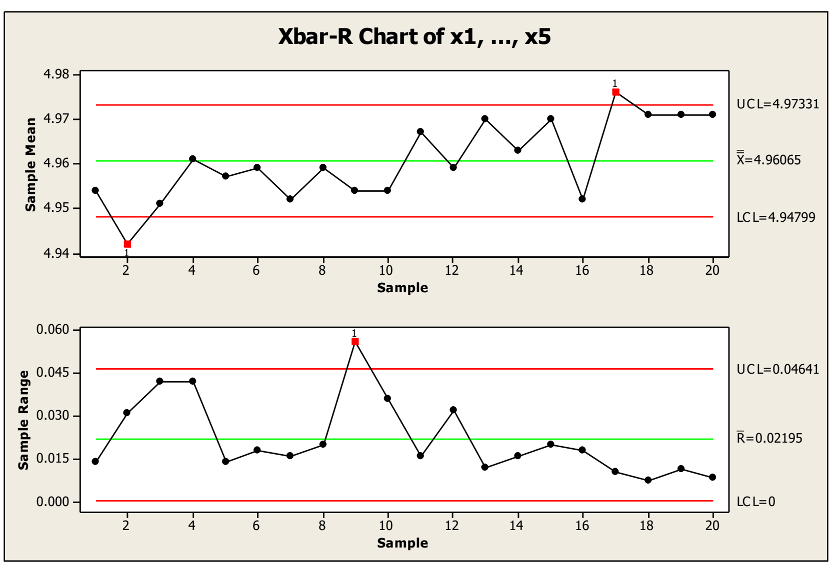


15-10 The following data were considered in *Quality Engineering* ["An SPC Case Study on Stabilizing Syringe Lengths" (1999–2000, Vol. 12(1))]. The syringe length is measured during a pharmaceutical manufacturing process. The following table provides data (in inches) for 20 samples each of size 5.

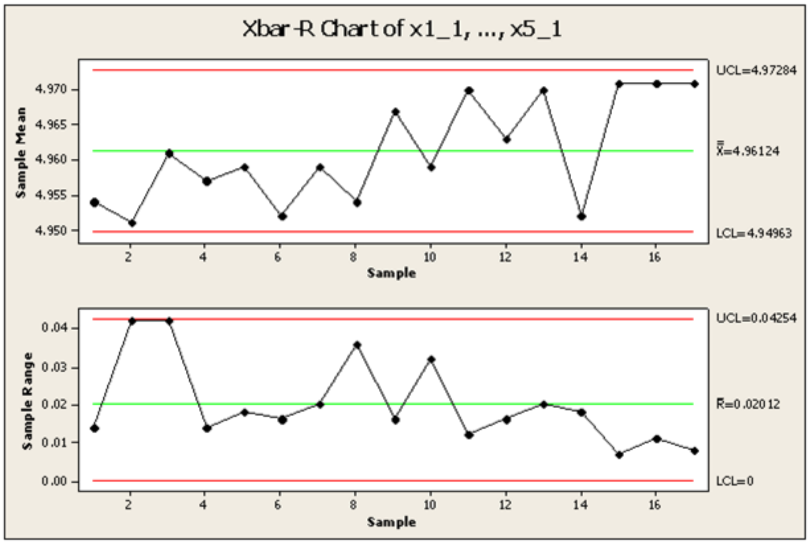
Sample	x_1	x_2	x_3	x_4	x_5
1	4.960	4.946	4.950	4.956	4.958
2	4.958	4.927	4.935	4.940	4.950
3	4.971	4.929	4.965	4.952	4.938
4	4.940	4.982	4.970	4.953	4.960
5	4.964	4.950	4.953	4.962	4.956
6	4.969	4.951	4.955	4.966	4.954
7	4.960	4.944	4.957	4.948	4.951
8	4.969	4.949	4.963	4.952	4.962
9	4.984	4.928	4.960	4.943	4.955
10	4.970	4.934	4.961	4.940	4.965
11	4.975	4.959	4.962	4.971	4.968
12	4.945	4.977	4.950	4.969	4.954
13	4.976	4.964	4.970	4.968	4.972
14	4.970	4.954	4.964	4.959	4.968
15	4.982	4.962	4.968	4.975	4.963
16	4.961	4.943	4.950	4.949	4.957
17	4.980	4.970	4.975	4.978	4.977
18	4.975	4.968	4.971	4.969	4.972
19	4.977	4.966	4.969	4.973	4.970
20	4.975	4.967	4.969	4.972	4.972

(a) Using all the data, find trial control limits for $\overline{X}$ and R charts, construct the chart, and plot the data. Is this process in statistical control?
(b) Use the trial control limits from part (a) to identify out-of-control points. If necessary, revise your control limits assuming that any samples that plot outside the control limits can be eliminated.
(c) Repeat parts (a) and (b) for $\overline{X}$ and S charts.

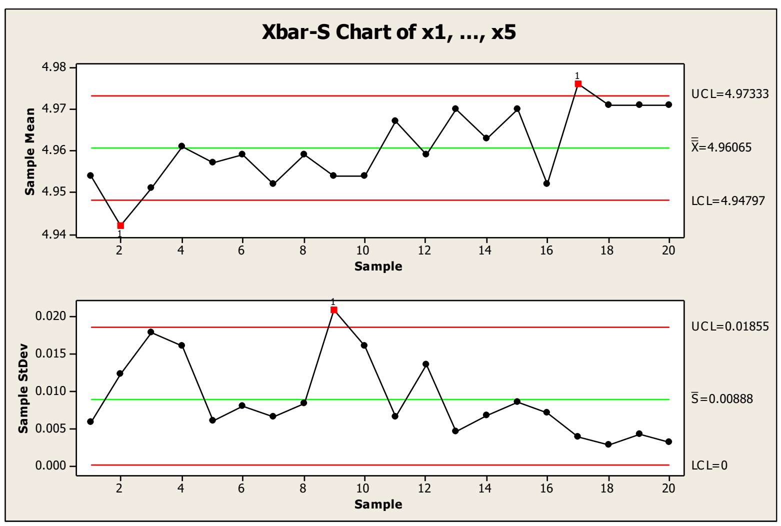
(a) The average range is used to estimate the standard deviation. Samples 2, 9, and 17 are out-of-control.

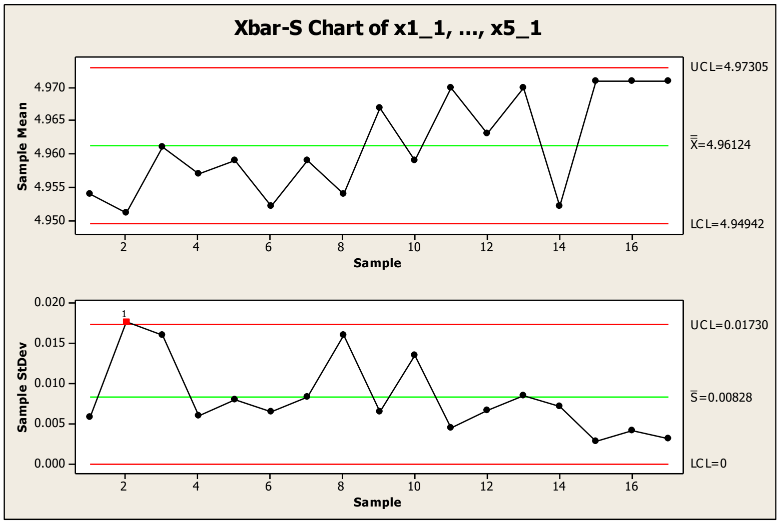


(b)



(c)For $\overline{X}$ -S chart



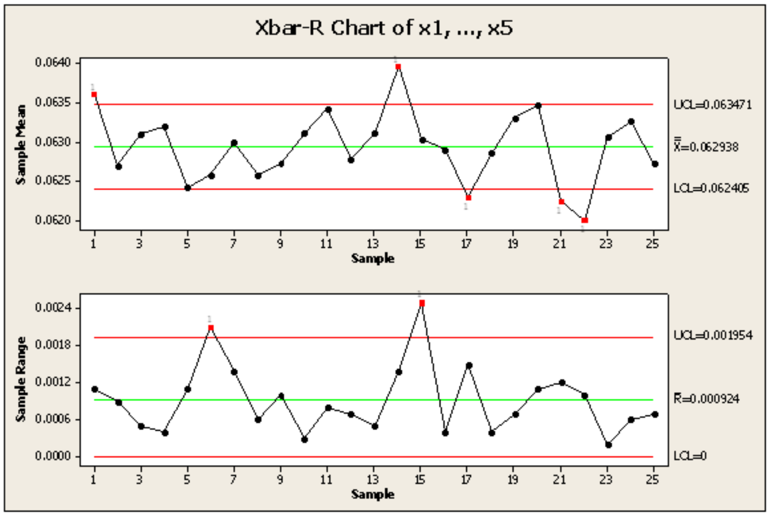


15-11 The thickness of a metal part is an important quality parameter. Data on thickness (in inches) are given in the following table, for 25 samples of five parts each.

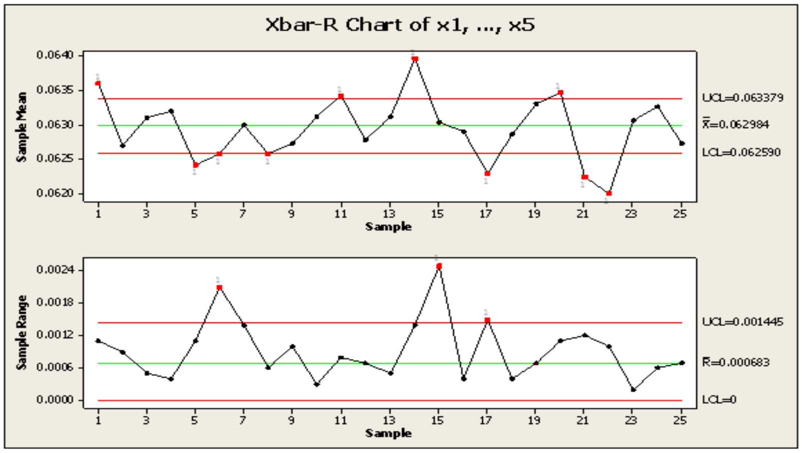
Sample Number	x_1	x_2	x_3	x_4	x_5
1	0.0629	0.0636	0.0640	0.0635	0.0640
2	0.0630	0.0631	0.0622	0.0625	0.0627
3	0.0628	0.0631	0.0633	0.0633	0.0630
4	0.0634	0.0630	0.0631	0.0632	0.0633
5	0.0619	0.0628	0.0630	0.0619	0.0625
6	0.0613	0.0629	0.0634	0.0625	0.0628
7	0.0630	0.0639	0.0625	0.0629	0.0627
8	0.0628	0.0627	0.0622	0.0625	0.0627
9	0.0623	0.0626	0.0633	0.0630	0.0624
10	0.0631	0.0631	0.0633	0.0631	0.0630
11	0.0635	0.0630	0.0638	0.0635	0.0633
12	0.0623	0.0630	0.0630	0.0627	0.0629
13	0.0635	0.0631	0.0630	0.0630	0.0630
14	0.0645	0.0640	0.0631	0.0640	0.0642
15	0.0619	0.0644	0.0632	0.0622	0.0635
16	0.0631	0.0627	0.0630	0.0628	0.0629
17	0.0616	0.0623	0.0631	0.0620	0.0625
18	0.0630	0.0630	0.0626	0.0629	0.0628
19	0.0636	0.0631	0.0629	0.0635	0.0634
20	0.0640	0.0635	0.0629	0.0635	0.0634
21	0.0628	0.0625	0.0616	0.0620	0.0623
22	0.0615	0.0625	0.0619	0.0619	0.0622
23	0.0630	0.0632	0.0630	0.0631	0.0630
24	0.0635	0.0629	0.0635	0.0631	0.0633
25	0.0623	0.0629	0.0630	0.0626	0.0628

(a) Using all the data, find trial control limits for $\bar{X}$ and R charts, construct the chart, and plot the data. Is this process in statistical control?
(b) Use the trial control limits from part (a) to identify out-of-control points. If necessary, revise your control limits assuming that any samples that plot outside the control limits can be eliminated.
(c) Repeat parts (a) and (b) for $\bar{X}$ and S charts.

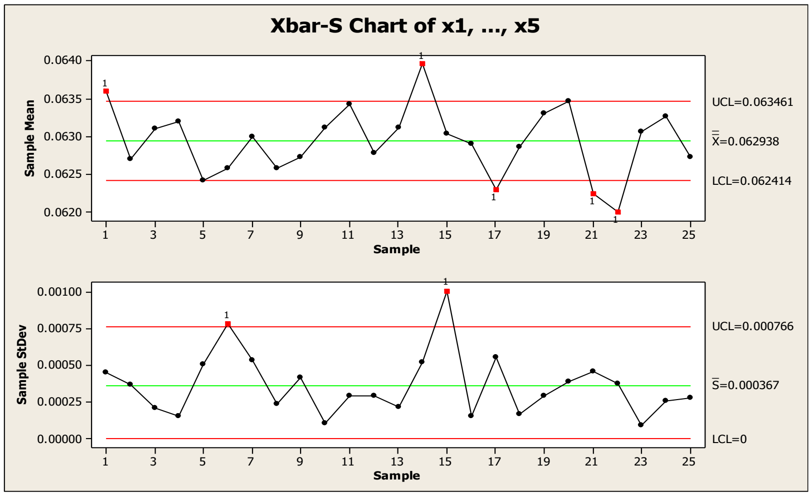
(a) The control limits for the following chart were obtained from $\bar{R}$.



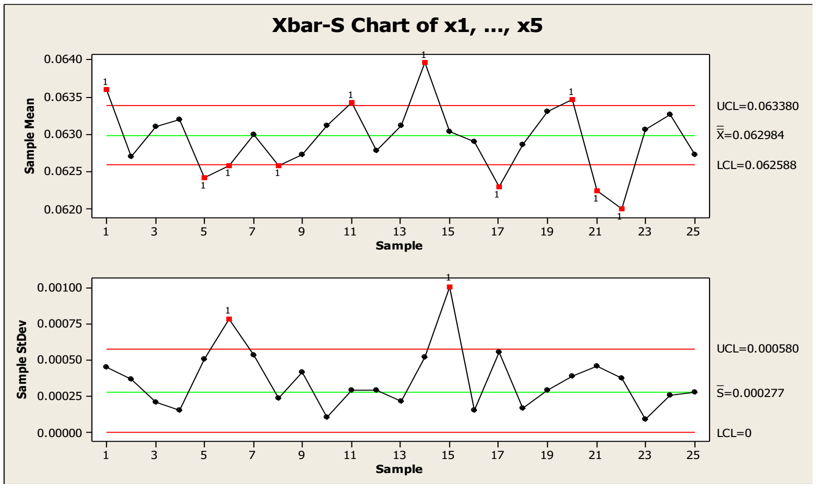
(b) The test failed at points 1, 6, 14, 15, 17, 21 and 22. The control limits are revised one time by omitting the out-of-control points. However, the charts still show additional out-of-control signals.



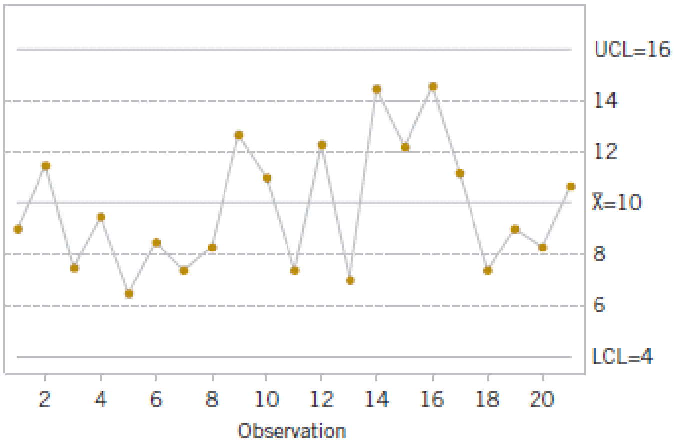
(c) The control limits for the following charts were obtained from $\bar{S}$.



(d)The test failed at points 1, 6, 14, 15, 17, 21 and 22. The control limits are revised one time by omitting the out-of-control points. However, the charts still show additional out-of-control signals.

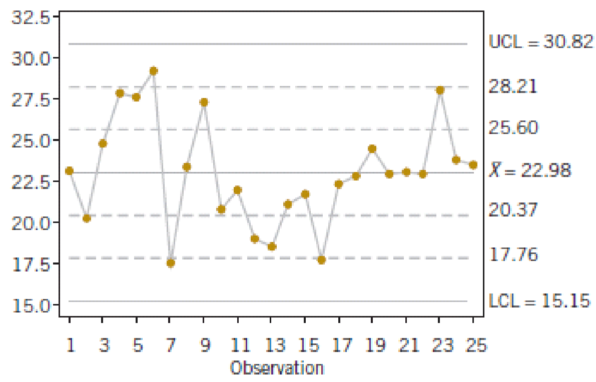


15-12 Apply the Western Electric Rules to the following $\bar{X}$ control chart. The warning limits are shown as dotted lines. Describe any rule violations.



It violates the rule of two out of three consecutive points plot beyond a 2-sigma limit.

15-13 Apply the Western Electric Rules to the following control chart. The warning limits are shown as dotted lines. Describe any rule violations.



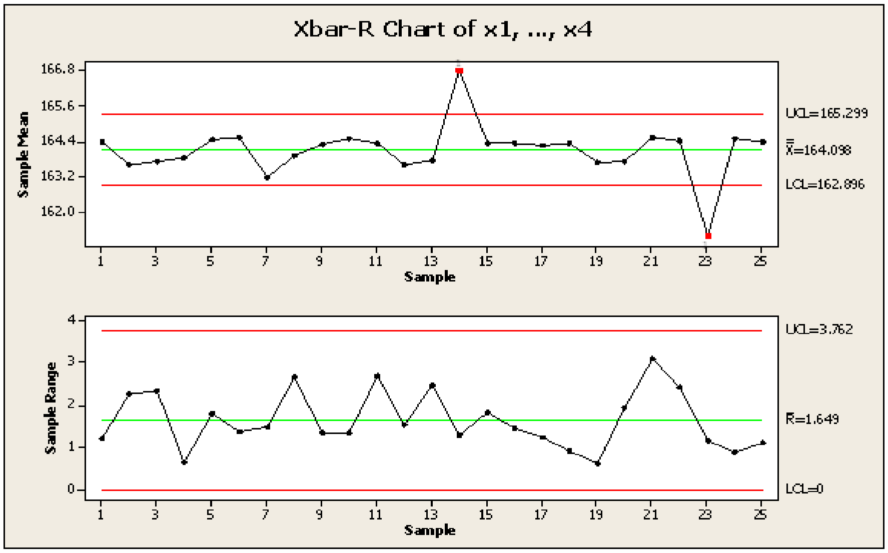
Points 3 to 7 fail the check for 4 out of 5 points beyond one sigma
Points 9 to 18 are 10 consecutive values below the center line

15-14 Web traffic can be measured to help highlight security problems or indicate a potential lack of bandwidth. Data on Web traffic (in thousand hits) from http://en.wikipedia.org/wiki/Web_traffic are given in the following table for 25 samples each of size 4.

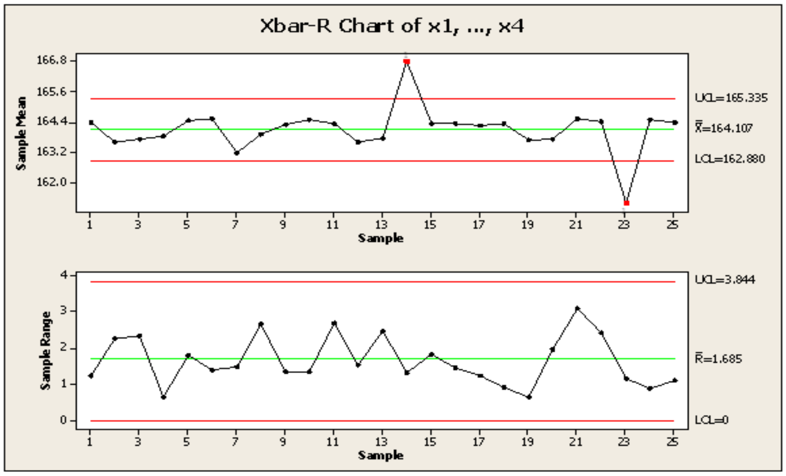
Sample	x_1	x_2	x_3	x_4
1	163.95	164.54	163.87	165.10
2	163.30	162.85	163.18	165.10
3	163.13	165.14	162.80	163.81
4	164.08	163.43	164.03	163.77
5	165.44	163.63	163.95	164.78
6	163.83	164.14	165.22	164.91
7	162.94	163.64	162.30	163.78
8	164.97	163.68	164.73	162.32
9	165.04	164.06	164.40	163.69
10	164.74	163.74	165.10	164.32
11	164.72	165.75	163.07	163.84
12	164.25	162.72	163.25	164.14
13	164.71	162.63	165.07	162.59
14	166.61	167.07	167.41	166.10
15	165.23	163.40	164.94	163.74
16	164.27	163.42	164.73	164.88
17	163.59	164.84	164.45	164.12
18	164.90	164.20	164.32	163.98
19	163.98	163.53	163.34	163.82
20	164.08	164.33	162.38	164.08
21	165.71	162.63	164.42	165.27
22	164.03	163.36	164.55	165.77
23	160.52	161.68	161.18	161.33
24	164.22	164.27	164.35	165.12
25	163.93	163.96	165.05	164.52

(a) Use all the data to determine trial control limits for $\bar{X}$ and R charts, construct the chart, and plot the data.
(b) Use the trial control limits from part (a) to identify out-of-control points. If necessary, revise your control limits, assuming that any samples that plot outside the control limits can be eliminated.

(a) The control limits for the following chart were obtained from $\overline{R}$.



(b) The test failed at points 14 and 23. The control limits are revised by omitting the out-of-control points from the control limit calculations.



An additional point is out-of-control and limits might be estimated again with this point eliminated.

15-15 Consider the data in Exercise 15-9. Calculate the sample standard deviation of all 60 measurements and compare this result to the estimate of σ obtained from your revised $\overline{X}$ and R charts. Explain any differences.

The sample standard deviation is 2.956. It is larger than the estimate obtained from the revised $\overline{X}$ and R chart, because the mean of the process has been shifted in the original data.

15-16 Consider the data in Exercise 15-10. Calculate the sample standard deviation of all 100 measurements and compare this result to the estimate of σ obtained from your revised $\overline{X}$ and R charts. Explain any differences.

The sample standard deviation is0.0127. It is larger than the estimate 0.010301 obtained from the revised $\overline{X}$ and R chart, because the mean of the process has been shifted in the original data.

15-17 An $\overline{X}$ control chart with 3-sigma control limits and subgroup size $n = 4$ has control limits $UCL = 48.75$ and $LCL = 40.55$.

(a) Estimate the process standard deviation.
(b) Does the response to part (a) depend on whether $\bar{r}$ or $\bar{s}$ was used to construct the $\bar{X}$ control chart?

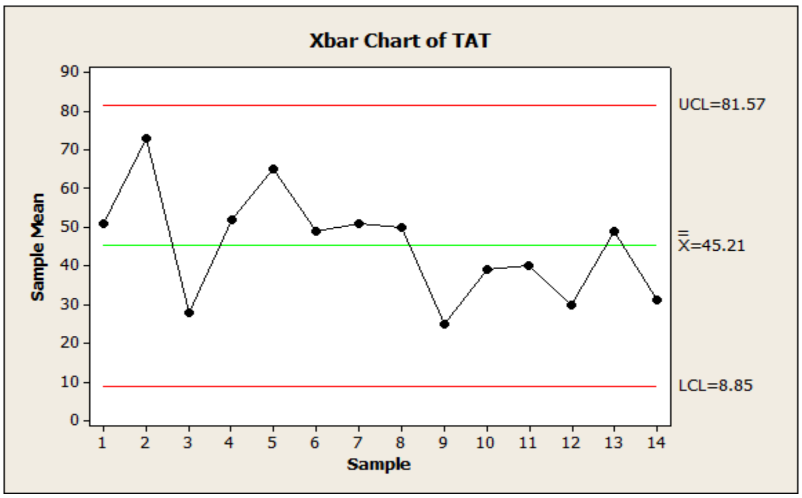
(a) The difference $UCL - LCL = 6\hat{\sigma}_{\bar{X}} = 48.75 - 40.55 = 8.2$
Therefore, $\hat{\sigma}_{\bar{X}} = \frac{8.2}{6} = 1.367$ because the limits are six standard errors wide. Then $\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{4}}$. Therefore, $\hat{\sigma} = 1.367(2) = 2.73$

(b) No, the calculation in part (a) is valid regardless of the method used to construct the control chart.

15-18 An article in *Quality & Safety in Health Care* ["Statistical Process Control as a Tool for Research and Healthcare Improvement," (2003)Vol. 12, pp. 458–464] considered a number of control charts in healthcare. The following approximate data were used to construct $\bar{X} - S$ charts for the turnaround time (TAT) for complete blood counts (in minutes). The subgroup size is *n* = 3 per shift, and the mean standard deviation is 21. Construct the $\bar{X}$ chart and comment on the control of the process. If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
TAT	51	73	28	52	65	49	51	50	25	39	40	30	49	31

$\bar{X} - S$ chart, and the process is in control



Section 15-4

15-19 Twenty successive hardness measurements are made on a metal alloy, and the data are shown in the following table.

Observation	Hardness	Observation	Hardness
1	51	11	51
2	52	12	57
3	54	13	58
4	55	14	50
5	55	15	53

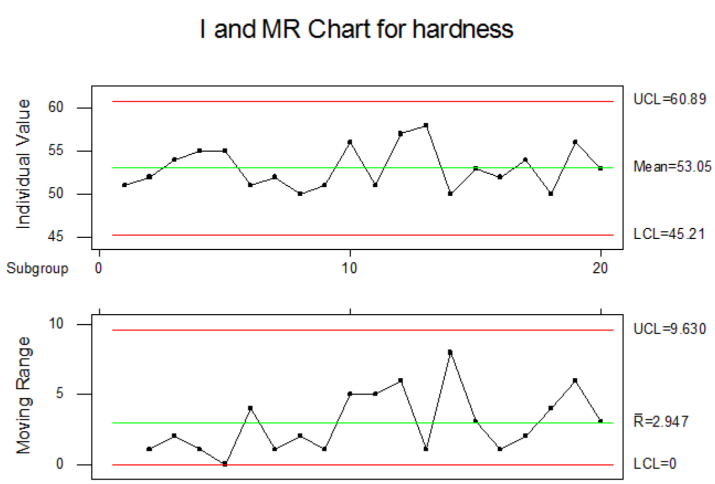
6	51	16	52
7	52	17	54
8	50	18	50
9	51	19	56
10	56	20	53

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.

(a)

```
                        Individuals and MR(2) - Initial Study
------------------------------------------------------------------------------
                             Charting Problem 15-17

Ind.x                                   | MR(2)
-----                                   | -----
UCL:  +     3.0 sigma = 60.8887         | UCL:  +     3.0 sigma = 9.63382
Centerline            = 53.05           | Centerline            = 2.94737
LCL:  -     3.0 sigma = 45.2113         | LCL:  -     3.0 sigma = 0
                                        |
out of limits = 0                       | out of limits = 0
------------------------------------------------------------------------------
Chart: Both      Normalize: No

   20 subgroups, size 1                      0 subgroups excluded

Estimated
process mean  = 53.05
process sigma =  2.61292
mean MR       =  2.94737
```

There are no points beyond the control limits. The process appears to be in control.



b) Estimated process mean and standard deviation

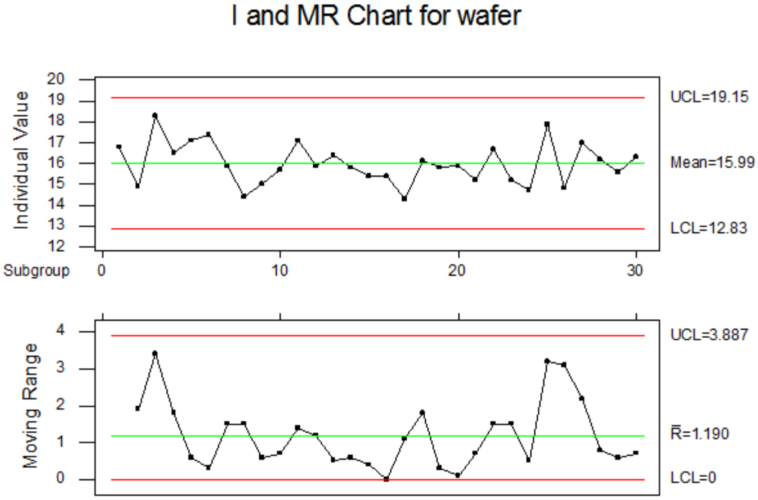
$$\hat{\mu} = \bar{x} = 53.05 \qquad \hat{\sigma} = \frac{\overline{mr}}{d_2} = \frac{2.94737}{1.128} = 2.613$$

15-20 In a semiconductor manufacturing process, CVD metal thickness was measured on 30 wafers obtained over approximately two weeks. Data are shown in the following table.

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.

Wafer	x	Wafer	x
1	16.8	16	15.4
2	14.9	17	14.3
3	18.3	18	16.1
4	16.5	19	15.8
5	17.1	20	15.9
6	17.4	21	15.2
7	15.9	22	16.7
8	14.4	23	15.2
9	15.0	24	14.7
10	15.7	25	17.9
11	17.1	26	14.8
12	15.9	27	17.0
13	16.4	28	16.2
14	15.8	29	15.6
15	15.4	30	16.3

(a) The process appears to be in statistical control. There are no points beyond the control limits.



(b) Estimated process mean and standard deviation

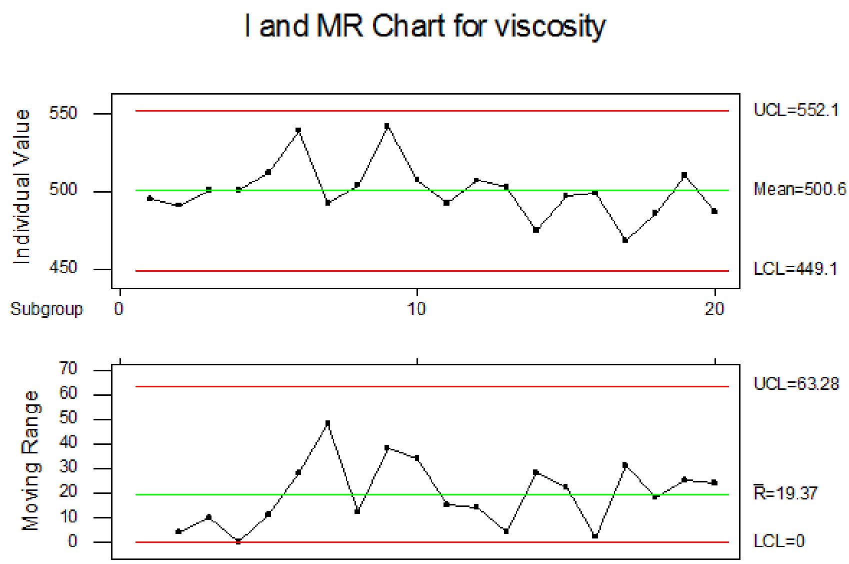
$$\hat{\mu} = \bar{x} = 15.99 \qquad \hat{\sigma} = \frac{\overline{mr}}{d_2} = \frac{1.190}{1.128} = 1.05$$

15-21 An automatic sensor measures the diameter of holes in consecutive order. The results of measuring 25 holes are in the following table.

Sample	Diameter	Sample	Diameter
1	9.94	14	9.99
2	9.93	15	10.12
3	10.09	16	9.81
4	9.98	17	9.73
5	10.11	18	10.14
6	9.99	19	9.96
7	10.11	20	10.06
8	9.84	21	10.11
9	9.82	22	9.95
10	10.38	23	9.92
11	9.99	24	10.09
12	10.41	25	9.85
13	10.36		

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the control chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.

(a)
```
                        Ind.x and MR(2) - Initial Study
--------------------------------------------------------------------------------
                            Charting Problem 15-20

Ind.x                                   | MR(2)
-----                                   | -----
UCL:  +    3.0 sigma = 552.112          | UCL:  +    3.0 sigma = 63.308
Centerline           = 500.6            | Centerline           = 19.3684
LCL:  -    3.0 sigma = 449.088          | LCL:  -    3.0 sigma =  0
                                        |
out of limits = 0                       | out of limits = 0
--------------------------------------------------------------------------------
Chart: Both      Normalize: No

   20 subgroups, size 1                       0 subgroups excluded
Estimated
process mean  = 500.6
process sigma =  17.1706
mean MR(2)    =  19.3684
```



(b) Estimated process mean and standard deviation

$$\hat{\mu} = \bar{x} = 500.6 \qquad \hat{\sigma} = \frac{\overline{mr}}{d_2} = \frac{19.3684}{1.128} = 17.17$$

15-22 The viscosity of a chemical intermediate is measured every hour. Twenty samples each of size $n = 1$ are in the following table.

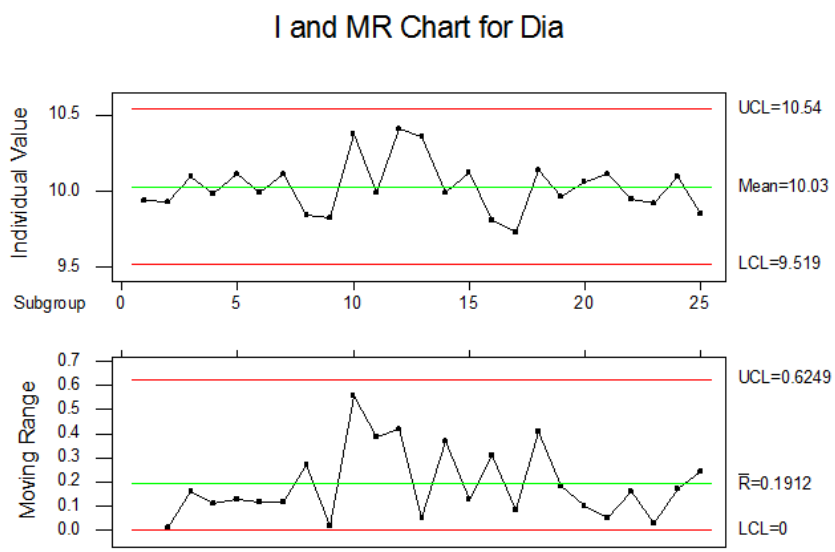
Sample	Viscosity	Sample	Viscosity
1	495	11	493
2	491	12	507
3	501	13	503
4	501	14	475
5	512	15	497
6	540	16	499
7	492	17	468
8	504	18	486
9	542	19	511
10	508	20	487

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.

(a) Ind.x and MR(2) - Initial Study

```
-------------------------------------------------------------------------------
                             Charting diameter

Ind.x                                       | MR(2)
-----                                       | -----
UCL:  +    3.0 sigma = 10.5358              | UCL:  +    3.0 sigma = 0.625123
Centerline           = 10.0272              | Centerline           = 0.19125
LCL:  -    3.0 sigma =  9.51856             | LCL:  -    3.0 sigma = 0
                                            |
```

```
out of limits = 0                        | out of limits = 0
------------------------------------------------------------------------------
Chart: Both      Normalize: No

   25 subgroups, size 1                        0 subgroups excluded

Estimated
process mean  = 10.0272
process sigma =   0.169548
mean MR(2)    =   0.19125
```



There are no points beyond the control limits. The process appears to be in control.

(b) Estimated process mean and standard deviation

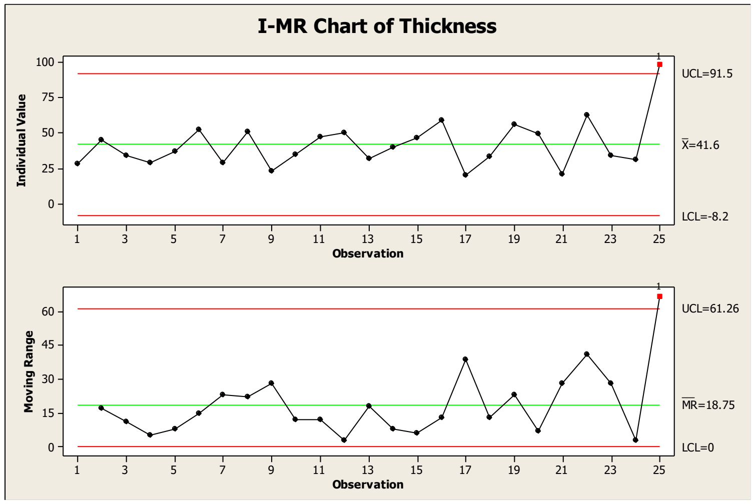
$$\hat{\mu} = \bar{x} = 10.0272 \qquad \hat{\sigma} = \frac{\overline{mr}}{d_2} = \frac{0.19125}{1.128} = 0.16955$$

15-23 The following table of data was analyzed in *Quality Engineering* [1991–1992, Vol. 4(1)]. The average particle size of raw material was obtained from 25 successive samples.

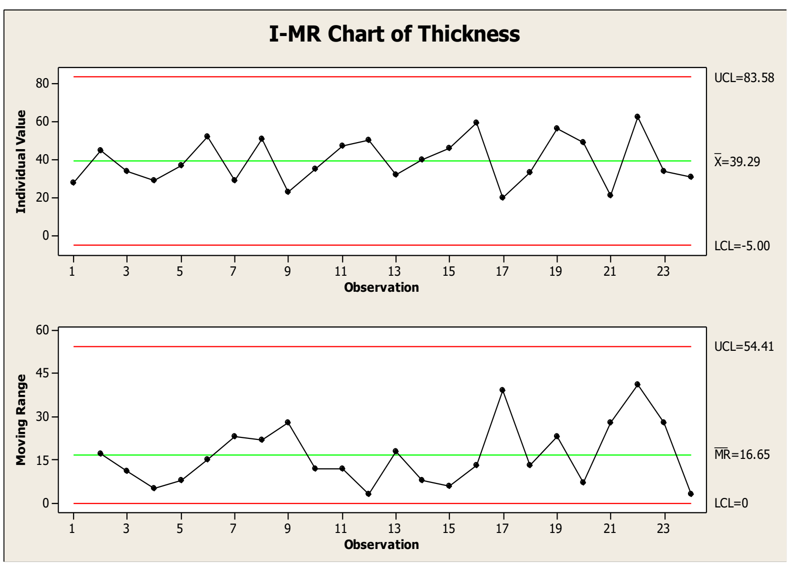
Observation	Size	Observation	Size
1	96.1	14	100.5
2	94.4	15	103.1
3	116.2	16	93.1
4	98.8	17	93.7
5	95.0	18	72.4
6	120.3	19	87.4
7	104.8	20	96.1
8	88.4	21	97.1
9	106.8	22	95.7
10	96.8	23	94.2
11	100.9	24	102.4
12	117.7	25	131.9
13	115.6		

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.

(a) The process is outof control. The control charts follow.



Remove the out-of-control observation 25:



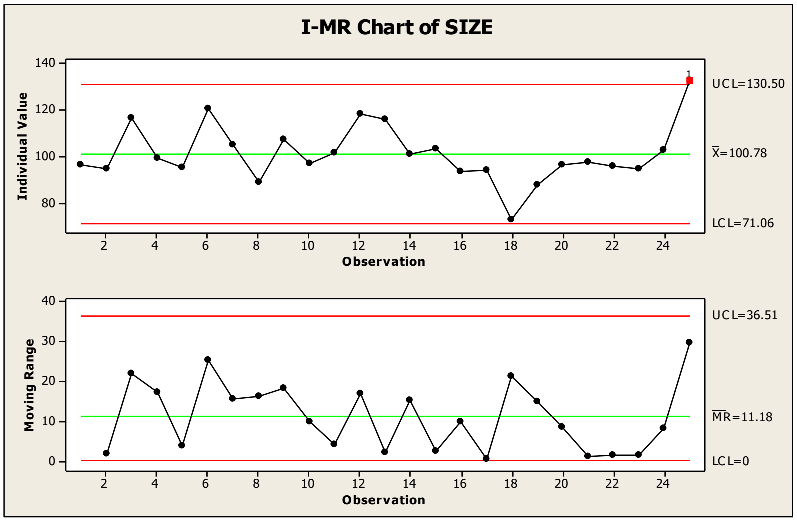
(b) From the centerline of the x chart, $\hat{\mu} = 39.29$. Also, $\hat{\sigma} = \overline{MR}/d_2$ = 16.65/1.128 = 14.76

15-24 Pulsed laser deposition technique is a thin film deposition technique with a high-powered laser beam. Twenty-five films were deposited through this technique. The thicknesses of the films obtained are shown in the following table.

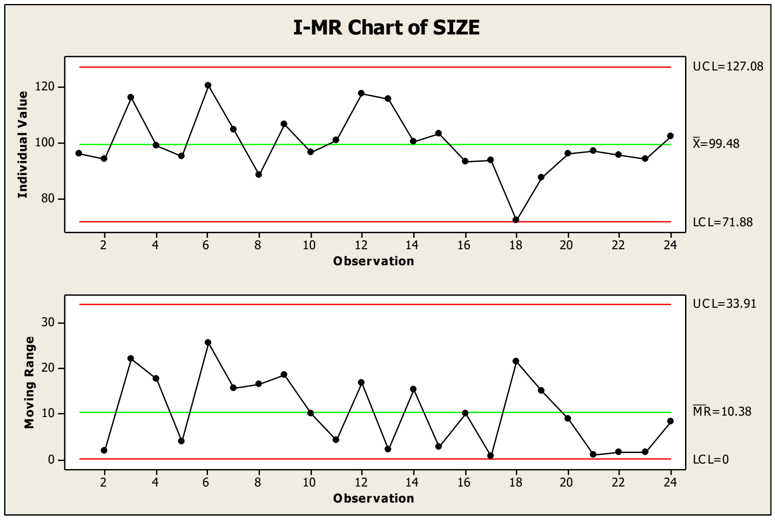
Film	Thickness (nm)	Film	Thickness (nm)
1	28	8	51
2	45	9	23
3	34	10	35
4	29	11	47
5	37	12	50
6	52	13	32
7	29	14	40
15	46	21	21
16	59	22	62
17	20	23	34
18	33	24	31
19	56	25	98
20	49		

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples, and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.

(a)



Remove the out-of-control observation:



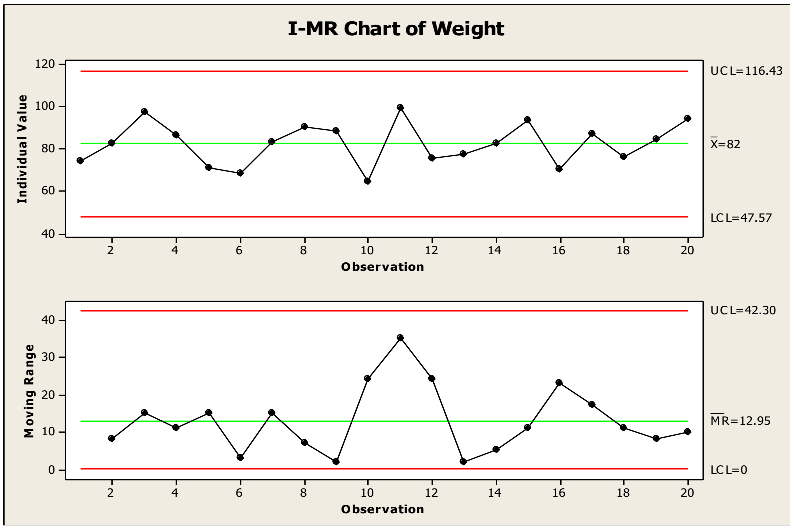
(b) The estimated mean is 99.4792 and estimated standard deviation is 9.20059.

15-25 The production manager of a soap manufacturing company wants to monitor the weights of the bars produced on the line. Twenty bars are taken during a stable period of the process. The weights of the bars are shown in the following table.

Bar	Weight (g)	Bar	Weight (g)
1	74	11	99
2	82	12	75
3	97	13	77
4	86	14	82
5	71	15	93
6	68	16	70
7	83	17	87
8	90	18	76
9	88	19	84
10	64	20	94

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples, and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.

(a) The process is in control. The control charts from Minitab follow.



(b) From the centerline of the x chart, $\hat{\mu} = 82$.
From x chart the difference $UCL - LCL = 116.43 - 47.57 = 68.86 = 6\hat{\sigma}$
Therefore, $\hat{\sigma} = 68.86/6 = 11.48$

15-26 An article in *Quality & Safety in Health Care* ["Statistical Process Control as a Tool for Research and Healthcare Improvement," (2003 Vol. 12, pp. 458–464)] considered a number of control charts in healthcare. An *X* chart was constructed for the amount of infectious waste discarded each day (in pounds). The article mentions that improperly classified infectious waste (actually not hazardous) adds substantial costs to hospitals each year. The following tables show approximate data for the average daily waste per month before and after process changes, respectively. The process change included an education campaign to provide an operational definition for infectious waste.

Before Process Change

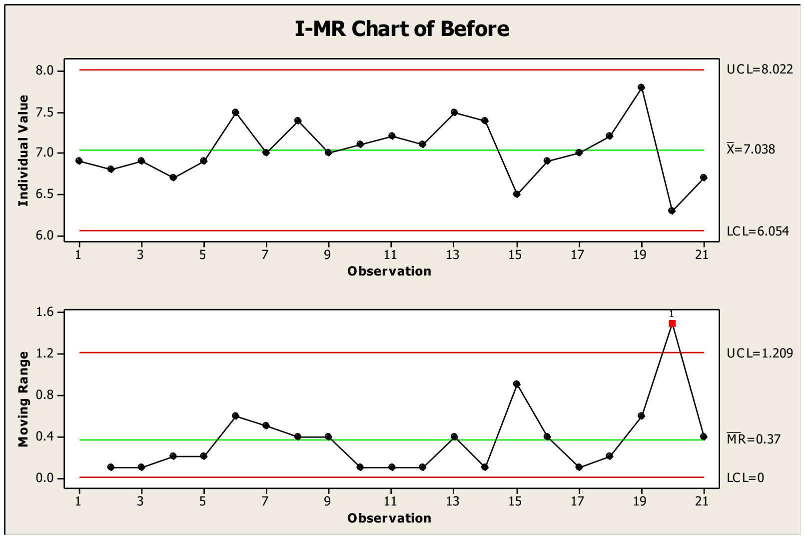
Month	1	2	3	4	5	6	7	8	9
Waste	6.9	6.8	6.9	6.7	6.9	7.5	7	7.4	7
Month	13	14	15	16	17	18	19	20	21
Waste	7.5	7.4	6.5	6.9	7.0	7.2	7.8	6.3	6.7

After Process Change

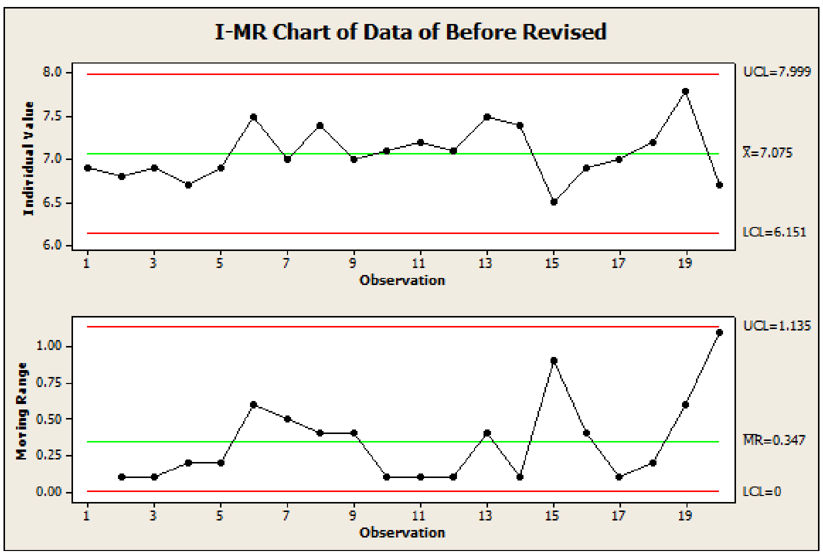
Month	1	2	3	4	5	6	7	8	9	10	11	12
Waste	5.0	4.8	4.4	4.3	4.6	4.3	4.5	3.5	4.0	4.1	3.8	5.0
Month	13	14	15	16	17	18	19	20	21	22	23	24
Waste	4.6	4.0	5.0	4.9	4.9	5.0	6.0	4.5	4.0	5.0	4.5	4.6
Month	25	26	27	28	29	30						
Waste	4.6	3.8	5.3	4.5	4.4	3.8						

(a) Handle the data before and after the process change separately and construct individuals and moving-range charts for each set of data. Assume that assignable causes can be found and eliminate suspect observations. If necessary, revise the control limits.
(b) Comment on the control of each chart and differences between the charts. Was the process change effective?

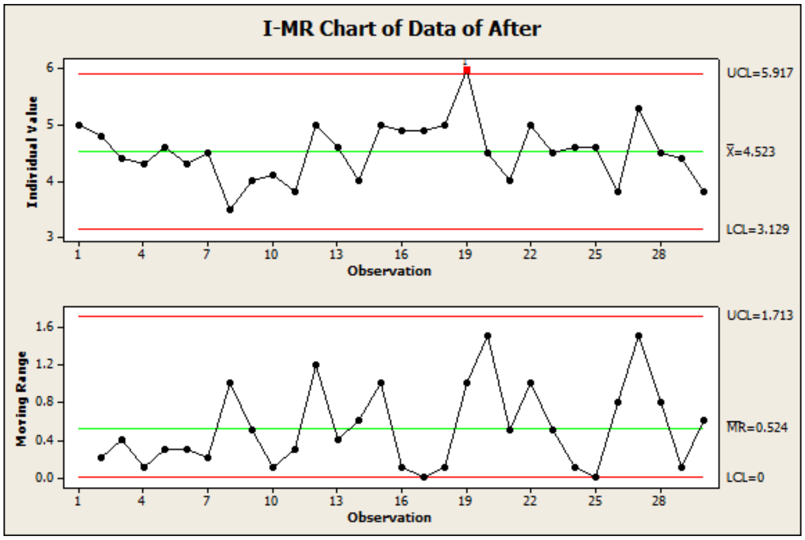
(a) For data before the process change, point 20 on the moving range chart is more than 3 standard deviations from center line.



Assume the assignable cause can be found and eliminate point 20 and revise the control chart.



For data of after process change, point 19 on the individuals chart is more than 30 standard deviations from the centerline.



Assume the assignable cause can be found and eliminate point 19 and revise the control chart.

(b) Consider a hypothesis test on the mean and standard deviation of data before and after the change (with point 20 removed from the before dataset and point 19 removed from after dataset)

Hypothesis test on the mean:

```
Two-sample T for Data Before Revised vs Data After Revised

                          N   Mean  StDev  SE Mean
Data of Before Revised   20  7.075  0.321    0.072
Data of After Revised    29  4.472  0.455    0.085

Difference = mu (Data of Before Revised) - mu (Data of After Revised)
Estimate for difference:  2.603
95% CI for difference:  (2.379, 2.826)
T-Test of difference = 0 (vs not =): T-Value = 23.47
P-Value = 0.000  DF = 46
```

```
Hypothesis test on the standard deviation
Method

Null        Sigma(Data Before Revised) / Sigma(Data After Revised) = 1
Alternative Sigma(Data Before Revised) / Sigma(Data After Revised) not = 1
Significance level      Alpha = 0.05

Statistics
Variable                 N  StDev  Variance
Data Before Revised     20  0.321     0.103
Data After Revised      29  0.455     0.207

Ratio of standard deviations = 0.705
Ratio of variances = 0.498


95% Confidence Intervals
                                CI for
Distribution   CI for StDev     Variance
of Data           Ratio           Ratio
Normal        (0.470, 1.095)  (0.221, 1.199)
Continuous    (0.419, 1.131)  (0.175, 1.279)


Tests
                                            Test
Method                        DF1  DF2  Statistic  P-Value
F Test (normal)                19   28       0.50    0.117
Levene's Test (any continuous)  1   47       2.43    0.126
```

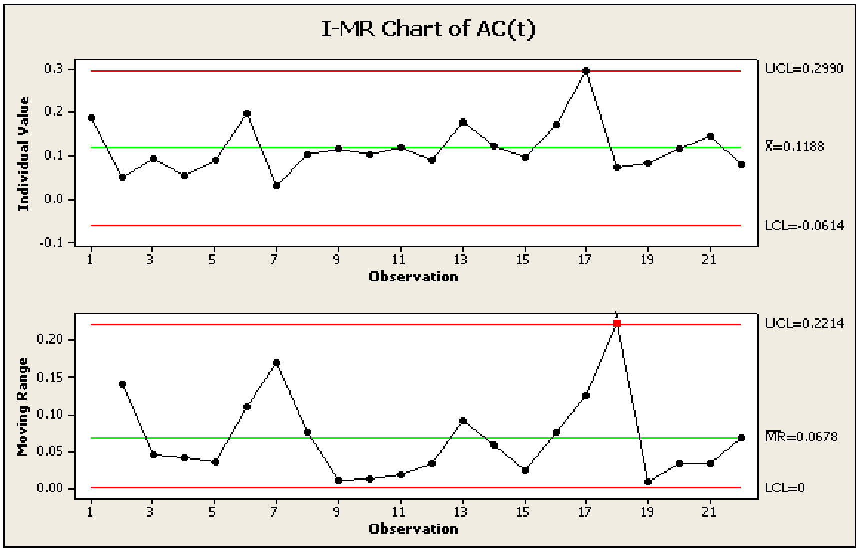
There is a significant change in the mean, but no significant change in the standard deviation.

15-27 An article in *Journal of the Operational Research Society* ["A Quality Control Approach for Monitoring Inventory Stock Levels," (1993, pp. 1115–1127)] reported on a control chart to monitor the accuracy of an inventory management system. Inventory accuracy at time t, $AC(t)$, is defined as the difference between the recorded and actual inventory (in absolute value) divided by the recorded inventory. Consequently, $AC(t)$ ranges between 0 and 1 with lower values better. Extracted data are shown in the following table.

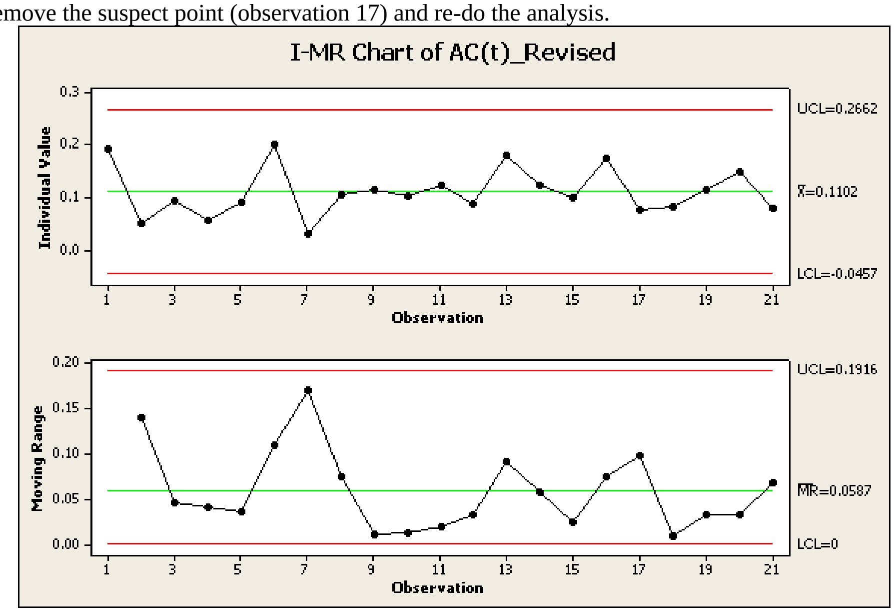
(a) Calculate individuals and moving-range charts for these data.
(b) Comment on the control of the process. If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.

t	1	2	3	4	5	6	7	8	9	10	11
$AC(t)$	0.190	0.050	0.095	0.055	0.090	0.200	0.030	0.105	0.115	0.103	0.121
t	12	13	14	15	16	17	18	19	20	21	22
$AC(t)$	0.089	0.180	0.122	0.098	0.173	0.298	0.075	0.083	0.115	0.147	0.079

(a) Individuals and moving range charts follow



(b) There is an out of control point on the moving range chart (|observation 18-observation 17| with moving range 0.223). Remove the suspect point (observation 17) and re-do the analysis.



<u>Section 15-5</u>

15-28 Suppose that a quality characteristic is normally distributed with specifications at 100 ± 20. The process standard deviation is 6.

(a) Suppose that the process mean is 100. What are the natural tolerance limits? What is the fraction defective? Calculate *PCR* and PCR_k and interpret these ratios.
(b) Suppose that the process mean is 106. What are the natural tolerance limits? What is the fraction defective? Calculate *PCR* and PCR_k and interpret these ratios.

(a)

The natural tolerance limits are 100 ± 18.

$$PCR = \frac{USL - LSL}{6\sigma} = \frac{40}{36} = 1.1111$$

Since the mean of the process is centered at the nominal dimension,

$$PCR_k = PCR = 1.1111$$

Since the process natural tolerance limits lie inside the specifications, very few defective units will be produced.

The fraction defective is $2\Phi(-20/6) = 0.0858\%$

(b)

The natural tolerance limits are 106 ± 18.

$$PCR = \frac{USL - LSL}{6\sigma} = \frac{40}{36} = 1.1111$$

Since the mean of the process is not centered at the nominal dimension,

$$PCR_k = \min\left[\frac{14}{18}, \frac{26}{18}\right] = 0.7778$$

The small PCR_K indicates that the process is likely to produce units outside the specification limits.
The fraction defective is

$$P(X < LSL) + P(X > USL) = P\left(Z < -14/6\right) + P\left(Z > 26/6\right) = 0.0098+0=0.0098$$

15-29 Suppose that a quality characteristic is normally distributed with specifi cations from 20 to 32 units.

(a) What value is needed for s to achieve a *PCR* of 1.5?
(b) What value for the process mean minimizes the fraction defective? Does this choice for the mean depend on the value of σ*?*

(a) $PCR = \frac{USL - LSL}{6\sigma} = \frac{12}{6\sigma} = 1.5$ so $\sigma = 1.3333$

(b) (20+32)/2=26 When the process is centered at the nominal dimension, the fraction defective is minimized for anyσ.

15-30 Suppose that a quality characteristic is normally distributed with specifications from 10 to 30 units. The process standard deviation is 2 units.

(a) Calculate the natural tolerance limits, fraction defective, *PCR*, and PCR_k when the process mean is 20.
(b) Suppose that the process mean shifts higher by 1.5 standard deviations. Recalculate the quantities in part (b).
(c) Compare the results in parts (a) and (b) and comment on any differences.

(a) The natural tolerance limits are 20 ± 6.

$$PCR = \frac{USL - LSL}{6\sigma} = \frac{20}{12} = 1.6667$$

Because the mean of the process is centered at the nominal dimension,

$$PCR_k = PCR = 1.6667$$

Because the process natural tolerance limits lie inside the specifications, very few defective units will be produced.
The fraction defective is $2\Phi(-10/2) = 0$

(b) The natural tolerance limits is 23 ± 6.

$$PCR = \frac{USL - LSL}{6\sigma} = \frac{20}{12} = 1.6667$$

Since the mean of the process is not centered at the nominal dimension,

$$PCR_k = \min\left[\frac{13}{6}, \frac{7}{6}\right] = 1.1667$$

The fraction defective is

$$P(X < LSL) + P(X > USL) = P(Z < -13/2) + P(Z > 7/2) = 0 + 0.00023 = 0.00023$$

(c) The measure of actual capability decreases and the fraction defective increases when the process mean is shifted from the center of the specification limits.

15-31 A normally distributed process uses 66.7% of the specification band. It is centered at the nominal dimension, located halfway between the upper and lower specification limits.

(a) Estimate *PCR* and PCR_k. Interpret these ratios.
(b) What fallout level (fraction defective) is produced?

(a) If the process uses 66.7% of the specification band, then 6σ = 0.667(USL-LSL).
Because the process is centered

$$3\sigma = 0.667(USL-\mu) = 0.677(\mu - LSL) = 0.667(USL-\mu)$$

$$4.5\sigma = USL-\mu = \mu - LSU$$

$$PC = PCR_K \quad = \min\left[\frac{4.5\sigma}{3\sigma}, \frac{4.5\sigma}{3\sigma}\right] = 1.5$$

Because PCR and PCR_k exceed unity, the natural tolerance limits are inside the specification limits and few defective units should be produced.

(b) Assuming a normal distribution with 6σ = 0.667(USL − LSL) and a centered process, then
3σ = 0.667(USL −μ). Consequently, USL −μ = 4.5σ and μ– LSL = 4.5σ

$$P(X > USL) = P\left(Z > \frac{4.5\sigma}{\sigma}\right) = P(Z > 4.5) = 1 - P(Z < 4.5)$$
$$= 1 - 1 = 0$$

By symmetry, the fraction defective is 2[P(X > USL)] = 0.

15-32 A normally distributed process uses 85% of the specification band. It is centered at the nominal dimension, located halfway between the upper and lower specification limits.

(a) Estimate *PCR* and PCR_k. Interpret these ratios.
(b) What fallout level (fraction defective) is produced?

(a) If the process uses 85% of the spec band then 6σ = 0.85(USL − LSL) and

$$PCR = \frac{USL-LSL}{0.85(USL-LSL)} = \frac{1}{0.85} = 1.18$$

Then 3σ = 0.85(USL −□) = 0.85(μ– LSL)
Therefore,

$$PCR_k = \min\left[\frac{3.53\sigma}{3\sigma}, \frac{3.53\sigma}{3\sigma}\right] = 1.18$$

Because PCR and PCR_k exceed unity, the natural tolerance limits are inside the specification limits and few defective units should be produced.

(b) Assuming a normal distribution with 6σ = 0.85(USL − LSL) and a centered process, then
3σ = 0.85(USL −μ). Consequently, USL −μ = 3.5σ and μ– LSL = 3.5σ

$$P(X > USL) = P\left(Z > \frac{3.5\sigma}{\sigma}\right) = P(Z > 3.5) = 1 - P(Z < 3.5)$$
$$= 1 - 0.999767 = 0.000233$$

By symmetry, the fraction defective is 2[P(X > USL)] = 0.00046

15-33 Reconsider Exercise 15-1. Suppose that the quality characteristic is normally distributed with specification at 220 ± 40. What is the fallout level? Estimate PCR and PCR_k and interpret these ratios.

Assume a normal distribution with $\hat{\mu} = 223$ and $\hat{\sigma} = \dfrac{34.286}{2.326} = 14.74$

$$P(X < LSL) = P\left(Z < \frac{LSL - \hat{\mu}}{\hat{\sigma}}\right) = P\left(Z < \frac{180 - 223}{14.74}\right) = P(Z < -2.92)$$
$$= 0.00175$$
$$P(X > USL) = P\left(Z > \frac{USL - \hat{\mu}}{\hat{\sigma}}\right) = P\left(Z > \frac{260 - 223}{14.74}\right) = P(Z > 2.51)$$
$$= 1 - P(Z < 2.51) = 1 - 0.99396 = 0.00604$$

Therefore, the proportion nonconforming is given by
P(X<LSL) + P(X>USL) = 0.00175 + 0.00604 = 0.00779

$$PCR = \frac{USL - LSL}{6(\hat{\sigma})} = \frac{260 - 180}{6(14.74)} = 0.905$$
$$PCR_K = \min\left[\frac{USL - \bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}} - LSL}{3\hat{\sigma}}\right]$$
$$= \min\left[\frac{260 - 223}{3(14.74)}, \frac{223 - 180}{3(14.74)}\right]$$
$$= \min[0.837, 0.972]$$
$$= 0.837$$

The process capability is marginal.

15-34 Reconsider Exercise 15-2 in which the specification limits are 14.50 ± 0.50.

(a) What conclusions can you draw about the ability of the process to operate within these limits? Estimate the percentage of defective items that is produced.
(b) Estimate *PCR* and PCR_k. Interpret these ratios.

(a) Assume a normal distribution with $\hat{\mu} = 14.510$ and $\hat{\sigma} = \dfrac{\bar{r}}{d_2} = \dfrac{0.344}{2.326} = 0.148$

$$P(X < LSL) = P\left(Z < \frac{LSL - \hat{\mu}}{\hat{\sigma}}\right) = P\left(Z < \frac{14.00 - 14.51}{0.148}\right) = P(Z < -3.45)$$
$$= 0.00028$$
$$P(X > USL) = P\left(Z > \frac{USL - \hat{\mu}}{\hat{\sigma}}\right) = P\left(Z > \frac{15.00 - 14.51}{0.148}\right) = P(Z > 3.31)$$
$$= 1 - P(Z < 3.31) = 1 - 0.99953 = 0.00047$$

Therefore, the proportion nonconforming is given by
P(X<LSL) + P(X>USL) =0.00028 + 0.00047 = 0.00075

(b)

$$PCR = \frac{USL - LSL}{6(\hat{\sigma})} = \frac{15.00 - 14.00}{6(0.148)} = 1.13$$

$$\begin{aligned} PCR_K &= \min\left[\frac{USL - \bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}} - LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{15.00 - 14.51}{3(0.148)}, \frac{14.51 - 14.00}{3(0.148)}\right] \\ &= \min[1.104, 1.15] \\ &= 1.104 \end{aligned}$$

Because PCR and PCR_k exceed unity, the natural tolerance limits are inside the specification limits and few defective units should be produced.

Because $PCR_K \cong PCR$ the process appears to be centered.

15-35 Reconsider Exercise 15-3. Suppose that the variable is normally distributed with specifications at 220 ± 50. What is the proportion out of specifications? Estimate and interpret *PCR* and PCR_k.

(a) Assume a normal distribution with $\hat{\mu} = 223$ and $\hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{13.58}{0.9213} = 14.74$

$$P(X < LSL) = P\left(Z < \frac{LSL - \hat{\mu}}{\hat{\sigma}}\right) = P\left(Z < \frac{170 - 223}{14.74}\right) = P(Z < -3.60)$$
$$= 0.00016$$

$$P(X > USL) = P\left(Z > \frac{USL - \hat{\mu}}{\hat{\sigma}}\right) = P\left(Z > \frac{270 - 223}{14.75}\right) = P(Z > 3.18)$$
$$= 1 - P(Z < 3.18) = 1 - 0.99926 = 0.00074$$

Probability of producing a part outside the specification limits is 0.00016 + 0.00074 = 0.0009

$$PCR = \frac{USL - LSL}{6(\hat{\sigma})} = \frac{270 - 170}{6(14.74)} = 1.13$$

$$\begin{aligned} PCR_K &= \min\left[\frac{USL - \bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}} - LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{270 - 223}{3(14.75)}, \frac{223 - 170}{3(14.75)}\right] \\ &= \min[1.06, 1.19] \\ &= 1.06 \end{aligned}$$

Because PCR and PCR_k exceed unity, the natural tolerance limits are inside the specification limits and few defective units should be produced. The estimated proportion nonconforming is given by P(X < LSL) + P(X > USL) = 0.00016 + 0.00074 = 0.0009

15-36 Reconsider Exercise 15-4(a). Assuming that both charts exhibit statistical control and that the process specifications are at 20 ± 5, estimate *PCR* and PCR_k and interpret these ratios.

Assuming a normal distribution with $\hat{\mu}$ = 20.0 and $\hat{\sigma}$ = 1.4

$$PCR = \frac{USL - LSL}{6(\hat{\sigma})} = \frac{25-15}{6(1.4)} = 1.19$$

$$\begin{aligned} PCR_K &= \min\left[\frac{USL-\hat{\mu}}{3\hat{\sigma}}, \frac{\hat{\mu}-LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{25-20}{3(1.4)}, \frac{20-15}{3(1.4)}\right] \\ &= \min[1.19, 1.19] \\ &= 1.19 \end{aligned}$$

The process is capable.

15-37 Reconsider the diameter measurements in Exercise 15-7. Use the revised control limits and process estimates.

(a) Estimate *PCR* and PCR_k. Interpret these ratios.
(b) What percentage of defectives is being produced by this process?

(a) $\hat{\sigma} = \dfrac{\bar{r}}{d_2} = \dfrac{5.737}{2.326} = 2.446 \quad \hat{\sigma} = 0.0002466$

$$PCR = \frac{USL - LSL}{6(\hat{\sigma})} = \frac{0.5045-0.5025}{6(0.0002466)} = 1.35$$

$$\begin{aligned} PCR_K &= \min\left[\frac{USL-\bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}}-LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{0.5045-0.5034}{3(0.0002466)}, \frac{0.5034-0.5025}{3(0.0002466)}\right] \\ &= \min[1.489, 1.217] \\ &= 1.217 \end{aligned}$$

Because PCR and PCR_k exceed unity, the natural tolerance limits are inside the specification limits and few defective units should be produced.

Because $PCR_K \neq PCR$ the process is slightly off center.

(b) Assume a normal distribution with $\hat{\mu}$ = 0.5034 and $\hat{\sigma}$ = 0.0002466

$$P(X < LSL) = P\left(Z < \frac{LSL-\hat{\mu}}{\hat{\sigma}}\right) = P(Z < -3.65) = 0.00013$$

$$\begin{aligned} P(X > USL) &= P\left(Z > \frac{USL-\hat{\mu}}{\hat{\sigma}}\right) = P(Z > 4.46) = 1 - P(Z < 4.46) \\ &= 1 - 1 = 0 \end{aligned}$$

Therefore, the proportion nonconforming is given by

$P(X<LSL) + P(X>USL) = 0.00013 + 0 = 0.00013$

15-38 Reconsider the copper-content measurements in Exercise 15-8. Given that the specifications are at 6.0 ± 1.0, estimate *PCR* and PCR_k and interpret these ratios.

Assuming a normal distribution with $\hat{\mu}$ = 6.284 and $\hat{\sigma} = \dfrac{1.1328}{1.693} = 0.669$

$$PCR = \frac{USL - LSL}{6(\hat{\sigma})} = \frac{7-5}{6(0.669)} = 0.50$$

$$\begin{aligned} PCR_K &= \min\left[\frac{USL - \bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}} - LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{7-6.284}{3(0.669)}, \frac{6.284-5}{3(0.669)}\right] \\ &= \min\left[0.357, 0.640\right] \\ &= 0.357 \end{aligned}$$

The process capability is poor.

15-39 Reconsider the pull-strength measurements in Exercise 15-9. Estimate the fallout level if the specifications are 16 ± 5. Estimate *PCR* and interpret this ratio.

$$\hat{\sigma} = \frac{\bar{r}}{d_2} = \frac{2.25}{1.693} = 1.329$$

$$\bar{x} = 15.09$$

$$\begin{aligned} &P(X > 15) + P(X < 5) \\ &= P\left(Z > \frac{15-15.09}{1.693}\right) + P\left(Z < \frac{5-15.09}{1.693}\right) \\ &= P(Z > -0.053) + P(Z < -5.96) \\ &= 0.5120 + 0.0 = 0.5120 \end{aligned}$$

$$PCR = \frac{15-5}{6(1.693)} = 0.985$$

Because the estimated PCR is less than unity, the process capability is not good.

15-40 Reconsider the syringe lengths in Exercise 15-10. Suppose that the specifications are set at 4.90 and 5.00.

(a) Estimate the process standard deviation.
(b) Calculate *PCR* and PCR_k for the process.

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
x	100	0	4.9607	0.00127	0.0127	4.9270	4.9520	4.9620	4.9700

Variable	Maximum
x	4.9840

(a) From the center line of the revised S chart, sbar = 0.00888 and n = 5. Therefore, 0.00888 = $c_4\hat{\sigma}$ and c_4 = 0.94. Therefore, $\hat{\sigma} = \frac{0.00888}{0.94} = 0.00944$

(b) $PCR = \frac{USL-LSL}{6\sigma} = \frac{5-4.9}{6(0.00944)} = 1.76$

$$PCR_K = \min\left(\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right) = \min\left(\frac{5-4.96}{3(0.00944)}, \frac{4.96-4.9}{3(0.00944)}\right) = 1.41$$

15-41 Reconsider the hardness measurements in Exercise 15-19. Suppose that the specifications are 45 to 60.

(a) Estimate the process standard deviation.
(b) Calculate *PCR* and PCR_k for the process.

(a) From the referenced exercise, $\widehat{\sigma_X} = 2.613$
(b) $PCR = \frac{USL-LSL}{6\cdot\sigma_X} = \frac{60-45}{6(2.613)} = 0.96$

$$PCR_k = \min\left(\frac{USL-\bar{X}}{3\sigma_X}, \frac{\bar{X}-LSL}{3\sigma_X}\right) = \min\left(\frac{60-50.3}{3(2.613)}, \frac{50.3-45}{3(2.613)}\right) = 0.68$$

15-42 Reconsider the viscosity measurements in Exercise 15-22. The specifications are 500 ± 25. Calculate estimates of the process capability ratios *PCR* and PCR_k for this process and provide an interpretation.

Assuming a normal distribution with $\hat{\mu} = 500.6$ and $\hat{\sigma} = 17.17$

$$PCR = \frac{USL-LSL}{6(\hat{\sigma})} = \frac{525-475}{6(17.17)} = 0.49$$

$$\begin{aligned} PCR_K &= \min\left[\frac{USL-\bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}}-LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{525-500.6}{3(17.17)}, \frac{500.6-475}{3(17.17)}\right] \\ &= \min[0.474, 0.50] \\ &= 0.474 \end{aligned}$$

Because the process capability ratios are less than unity, the process capability appears to be poor.

15-43 Suppose that a quality characteristic is normally distributed with specifications at 120 ± 20. The process standard deviation is 6.5.

(a) Suppose that the process mean is 120. What are the natural tolerance limits? What is the fraction defective? Calculate *PCR* and PCR_k and interpret these ratios.
(b) Suppose that the process mean shifts off-center by 1.5 standard deviations toward the upper specification limit. Recalculate the quantities in part (a).
(c) Compare the results in parts (a) and (b) and comment on any differences.

(a) The natural tolerance limits are 120 ± 3(6.5) = (100.5, 139.5)
The fraction conforming is
$P(100 < X < 140) = P[(100 - 120)/6.5 < Z < (140 - 120)/6.5] = P[-3.0769 < Z < 3.0769] = 0.9979$
Therefore the fraction defective = 1 – 0.9979 = 0.002

PCR = 40/(6 × 6.5) = 1.03
PCR_k = 1.03 because the process is centered within the specifications.

(b) The shift is 1.5 × 6.5 = 9.75. The natural tolerance limits are 129.75 ± 3(6.5) = (110.25, 149.25)
The fraction conforming is
$P(100 < X < 140) = P[(100 - 129.75)/6.5 < Z < (140 - 129.75)/6.5] = P[-4.5769 < Z < 1.5769] = 0.9426$
Therefore, the fraction defective = 1 – 0.9426 = 0.057

PCR remains the same = 40/(6 × 6.5) = 1.03.
The nearest specification to the process mean is 140. Therefore,

$PCR_k = (140 - 129.75)/(3 \times 6.5) = 0.526$

(c) The fraction defective increases in part (b) when the process mean shifts from the center of the specifications. This change is reflected in the decreased PCR_k.

15-44 Suppose that a quality characteristic is normally distributed with specifications at 150 ± 20. Natural tolerance limits for the process are 150 ± 18.

(a) Calculate the process standard deviation.
(b) Calculate *PCR* and PCR_k of the process. Calculate the percentage of the specification width used by the process.
(c) What fallout level (fraction defective) is produced?

(a) The natural tolerance limits are $150 \pm 3(\sigma) = 150 \pm 18$
The standard deviation = 6.

(b) PCR = 40/ (6 × 6) = 1.11
PCR_k = 1.11 because the process is centered within the specifications.
The process width = 18 and the specification width = 20. The percentage of the specification width used by the process = 18/20 = 90%
(c) The fraction conforming is
$P(130 < X < 170) = P[(130 - 150)/6 < Z < (170 - 150)/6] = P[-3.33 < Z < 3.33] = 0.9991$
The fraction defective = 1 – the fraction conforming = 1- 0.9991 = 0.0009

15-45 An $\bar{X}$ control chart with 3-sigma control limits and subgroup size $n = 4$ has control limits $UCL = 28.8$ and $LCL = 24.6$. The process specification limits are (24, 32).

(a) Estimate the process standard deviation.
(b) Calculate *PCR* and PCR_k for the process.

(a) For the $\bar{x}$ chart: The difference $UCL - LCL = 6\hat{\sigma}_{\bar{X}} = 28.8 - 24.6 = 4.2$

Therefore, $\hat{\sigma}_{\bar{X}} = \dfrac{4.2}{6} = 0.7$ and $\hat{\sigma} = 0.7\sqrt{4} = 1.4$

(b) PCR = (32 – 24)/(6 × 1.4) = 0.9524
The control charts are centered at (28.8+24.6)/2 = 26.7 so this value estimates the process mean.
$PCR_k = (26.7 - 24)/(3 \times 1.4) = 0.6429$

15-46 A control chart for individual observations has 3-sigma control limits $UCL = 1.80$ and $LCL = 1.62$. The process specification limits are (1.64, 1.84).

(a) Estimate the process standard deviation.
(b) Calculate *PCR* and PCR_k for the process.

(a) The difference $UCL - LCL = 6\hat{\sigma} = 1.80 - 1.62 = 0.18$

Therefore, $\hat{\sigma} = \dfrac{0.18}{6} = 0.03$

(b) PCR = (1.84 – 1.64)/(6 × 0.03) = 1.11

The control charts are centered at (1.80 + 1.62)/2 = 1.71 so this value estimates the process mean.
$PCR_k = (1.71 - 1.64)/(3 \times 0.03) = 0.778$

15-47 A process mean is centered between the specification limits and $PCR = 1.33$. Assume that the process mean increases by 1.5σ.

(a) Calculate *PCR* and PCR_k for the shifted process.
(b) Calculate the estimated fallout from the shifted process and compare your result to those in Table 15-4. Assume a normal distribution for the measurement.

(a) There is no change to $PCR = \frac{USL-LSL}{6 \cdot \sigma_X} = 1.33$

USL – LSL = 1.33(6□) = 7.98□. Therefore, before the shift the distance from the process mean □ to the USL = 3.99□. After the shift, this distance = 2.49□. Therefore,

$$PCR_k = \frac{2.5\sigma}{3\sigma_X} = 0.83$$

(b) Before the shift, the distance from the process mean to the USL and to the LSL is 4s. After the process shift, the distances to USL and LSL are 2.5s and 5.5s, respectively. Therefore, the probability of fallout is P(Z > 2.49) + P(Z < 5.49) = 0.00639+ 2E-8 ≈ 0.00639 and this agrees with the result in the table.

15-48 The *PCR* for a measurement is 1.5 and the control limits for an $\bar{X}$ chart with *n* = 4 are 24.6 and 32.6.

(a) Estimate the process standard deviation σ.
(b) Assume that the specification limits are centered around the process mean. Calculate the specification limits.

(a) $6\hat{\sigma}_{\bar{X}} = UCL - LCL$, then $\hat{\sigma}_{\bar{X}} = \frac{32.6-24.6}{6} = \frac{4}{3}$

$\hat{\sigma} = \hat{\sigma}_{\bar{X}}\sqrt{4} = \frac{8}{3}$

(b) $PCR = \frac{USL-LSL}{6\sigma}$

$USL - LSL = PCR(6\hat{\sigma}) = 1.5(6)(\frac{8}{3}) = 24$. The process mean is centered between the control limits at 28.6.

Therefore, $USL = 28.6 + 12 = 40.6, LCL = 28.6 - 12 = 16.6$

$USL = 40.6, LSL = 16.6$

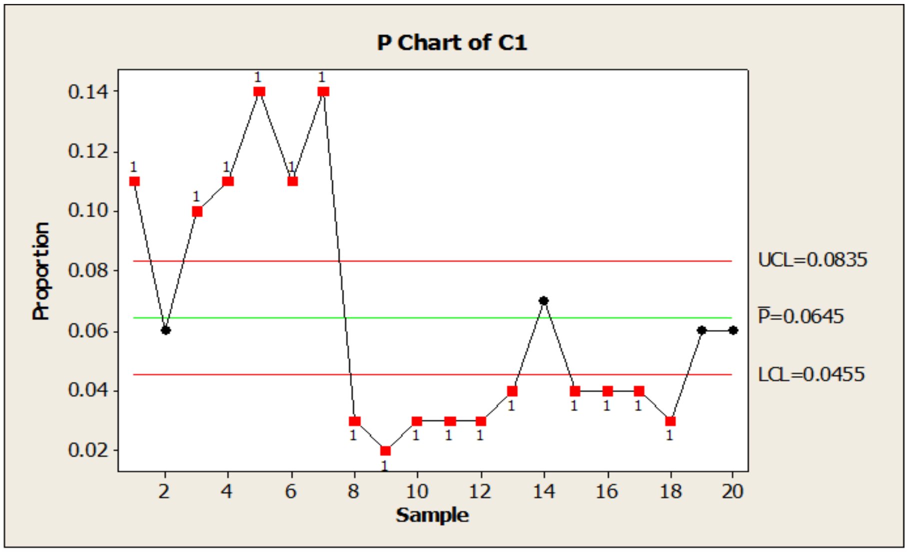
Section 15-6

15-49 An early example of SPC was described in *Industrial Quality Control* ["The Introduction of Quality Control at Colonial Radio Corporation" (1944, Vol. 1(1), pp. 4–9)]. The following are the fractions defective of shaft and washer assemblies during the month of April in samples of n = 1500 each:

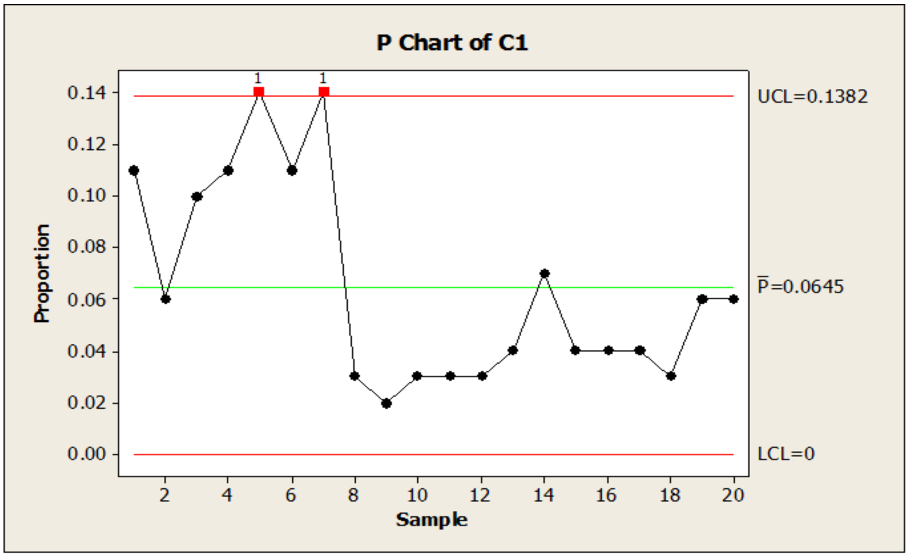
Sample	Fraction Defective	Sample	Fraction Defective
1	0.11	11	0.03
2	0.06	12	0.03
3	0.1	13	0.04
4	0.11	14	0.07
5	0.14	15	0.04
6	0.11	16	0.04
7	0.14	17	0.04
8	0.03	18	0.03
9	0.02	19	0.06
10	0.03	20	0.06

(a) Set up a P chart for this process. Is this process in statistical control?
(b) Suppose that instead of n = 1500, n = 100. Use the data given to set up a P chart for this process. Revise the control limits if necessary.
(c) Compare your control limits for the P charts in parts (a) and (b). Explain why they differ. Also, explain why your assessment about statistical control differs for the two sizes of n.

(a) This process is out of control

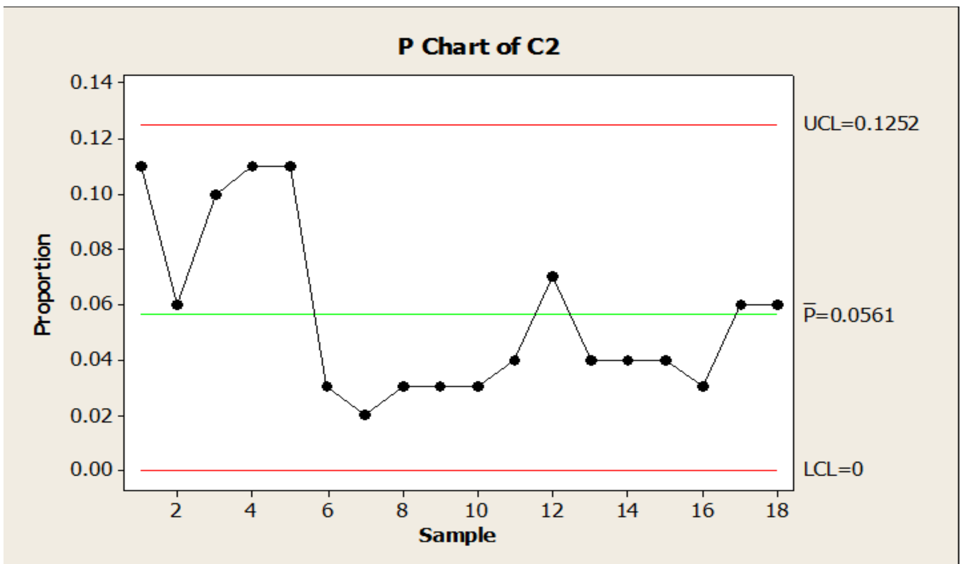


(b)



The process is still out of control, but not as many points fall outside of the control limits. The control limits are wider for smaller values of *n*.
Test Failed at points 5 and 7.

The following charteliminates points 5 and 7.



(c) The larger sample size leads to a smaller standard deviation for the proportions and.Thus. narrower control limits.

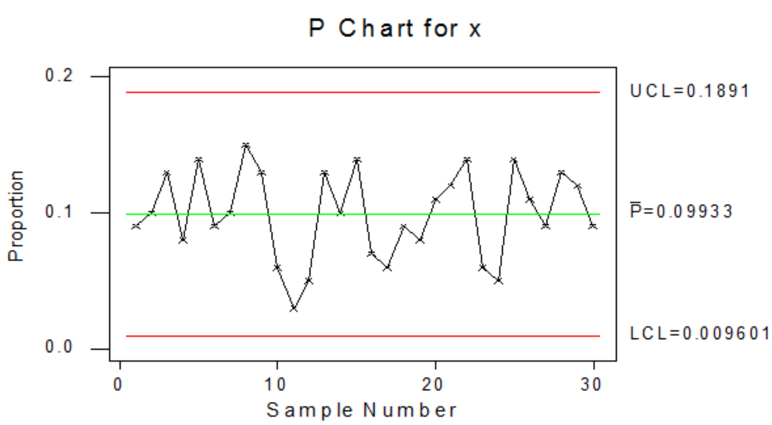
15-50 Suppose that the following fraction defective has been found in successive samples of size 100 (read down):

(a) Using all the data, compute trial control limits for a fraction-defective control chart, construct the chart, and plot the data.

0.09	0.03	0.12
0.10	0.05	0.14
0.13	0.13	0.06
0.08	0.10	0.05
0.14	0.14	0.14
0.09	0.07	0.11
0.10	0.06	0.09
0.15	0.09	0.13
0.13	0.08	0.12
0.06	0.11	0.09

(b) Determine whether the process is in statistical control. If not, assume that assignable causes can be found and out-of-control points eliminated. Revise the control limits.

(a)



(b) The process appears to be in statistical control.

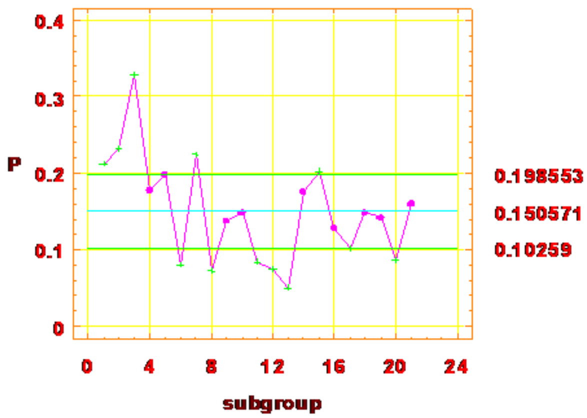
15-51 The following are the numbers of defective solder joints found during successive samples of 500 solder joints:

Day	No. of Defectives	Day	No. of Defectives
1	106	12	37
2	116	13	25
3	164	14	88
4	89	15	101
5	99	16	64
6	40	17	51
7	112	18	74
8	36	19	71
9	69	20	43
10	74	21	80
11	42		

(a) Using all the data, compute trial control limits for a fraction-defective control chart, construct the chart, and plot the

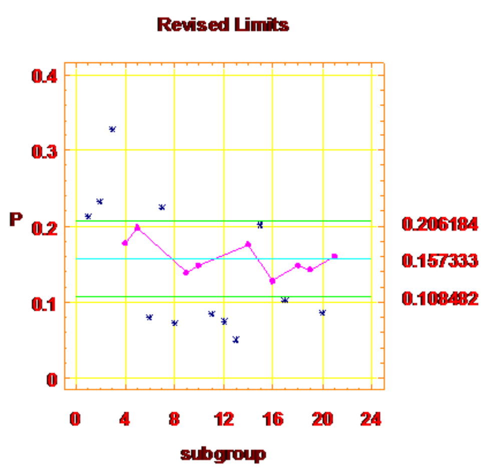
data.
(b) Determine whether the process is in statistical control. If not, assume that assignable causes can be found and out-of-control points eliminated. Revise the control limits.

```
                          P Chart - Initial Study
             Charting Problem 15-43
P Chart
-----
UCL:  +    3.0 sigma = 0.198553
Centerline           = 0.150571
LCL:  -    3.0 sigma = 0.10259


out of limits = 12
Estimated
mean P  = 0.150571
sigma   = 0.0159937
```



The samples with out-of-control points are 1, 2, 3, 6, 7, 8, 11, 12, 13, 15, 17, and 20. The control limits need to be revised.

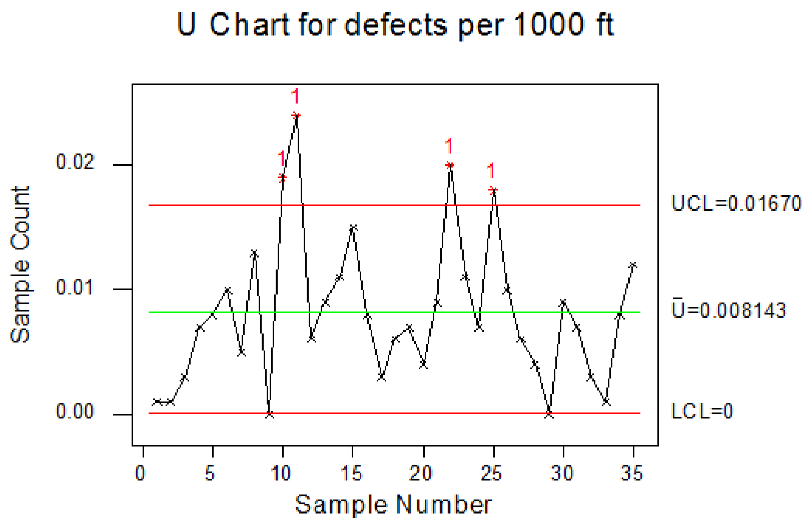
```
                     P Chart - Revised Limits
      Charting Problem 15-43
P Chart
-----
UCL:  +    3.0 sigma = 0.206184
Centerline           = 0.157333
LCL:  -    3.0 sigma = 0.108482
out of limits = 0
Estimated
mean P  = 0.157333
sigma   = 0.0162837
```



There are no further points out of control for the revised limits.

15-52 The following represent the number of defects per 1000 feet in rubber-covered wire: 1, 1, 3, 7, 8, 10, 5, 13, 0, 19, 24, 6, 9, 11, 15, 8, 3, 6, 7, 4, 9, 20, 11, 7, 18, 10, 6, 4, 0, 9, 7, 3, 1, 8, 12. Do the data come from a controlled process?

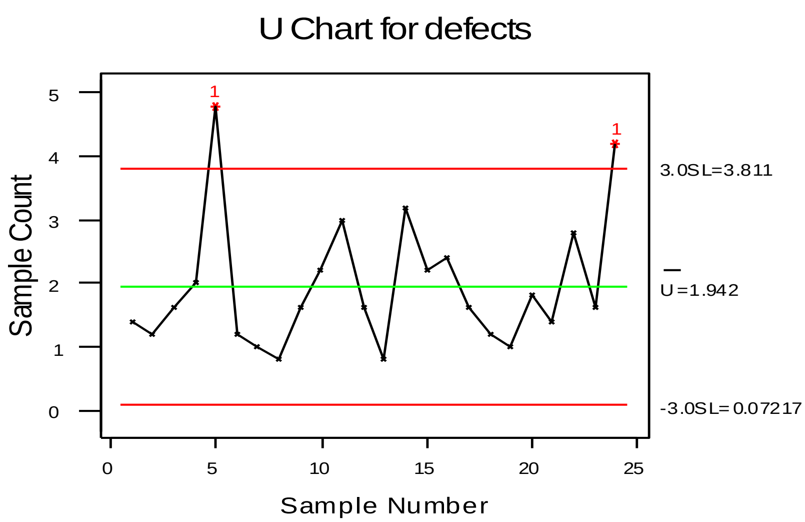
The process does not appear to be in control.



15-53 The following represent the number of solder defects observed on 24 samples of five printed circuit boards: 7, 6, 8, 10, 24, 6, 5, 4, 8, 11, 15, 8, 4, 16, 11, 12, 8, 6, 5, 9, 7, 14, 8, 21.

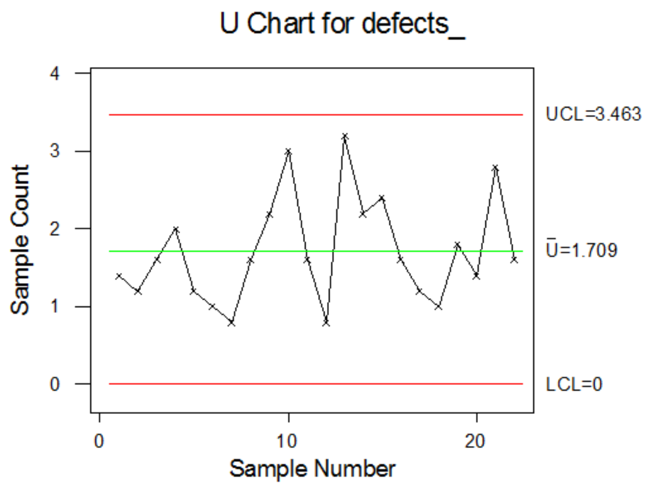
(a) Using all the data, compute trial control limits for a *U* control chart, construct the chart, and plot the data.
(b) Can we conclude that the process is in control using a *U* chart? If not, assume that assignable causes can be found, and list points and revise the control limits.

(a)



Samples 5 and 24 are points beyond the control limits. The limits need to be revised.

(b)

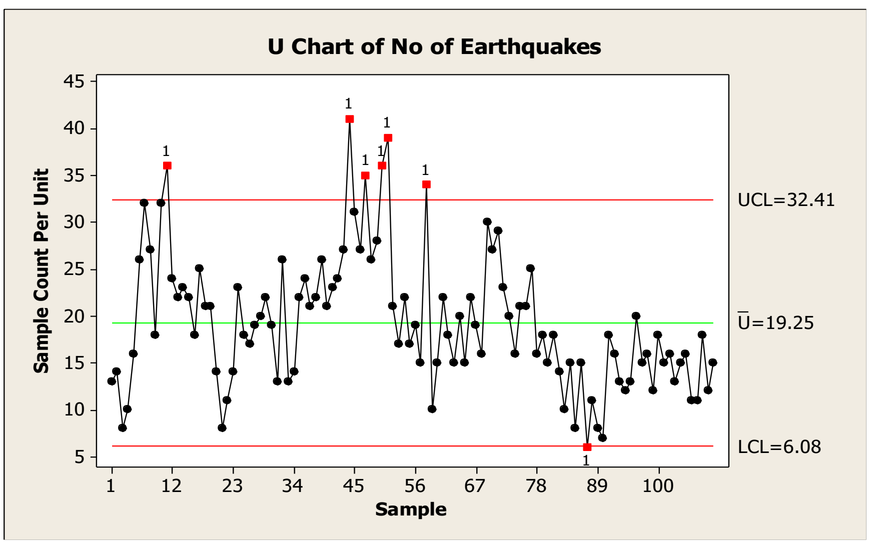


The control limits are calculated without the out-of-control points. There are no further points out of control for the revised limits.

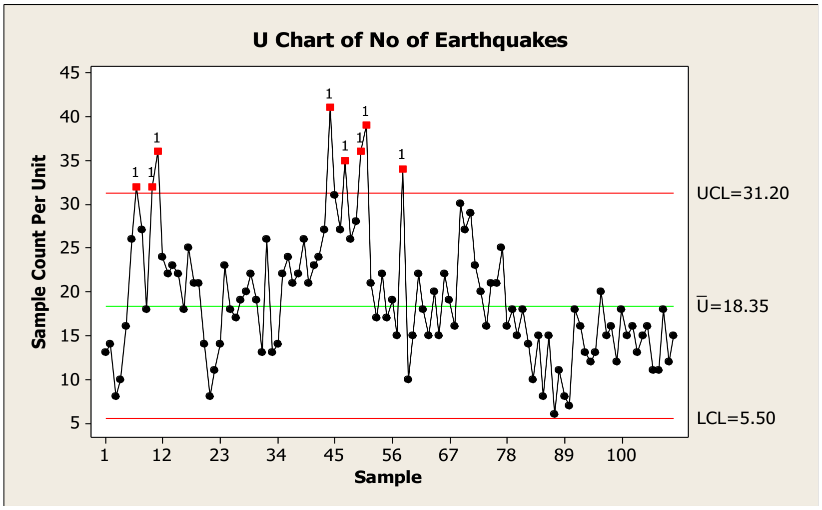
15-54 Consider the data on the number of earthquakes of magnitude 7.0 or greater by year in Exercise 6-87.

(a) Construct a *U* chart for this data with a sample size of $n = 1$.
(b) Do the data appear to be generated by an in-control process? Explain.

(a)



(b) No. The process is out-of-control at observations 11, 44, 47, 50, 51, 58, and 87. The control limits are revised one time by omitting the out-of-control points. However, the chart shows additional out-of-control signals.

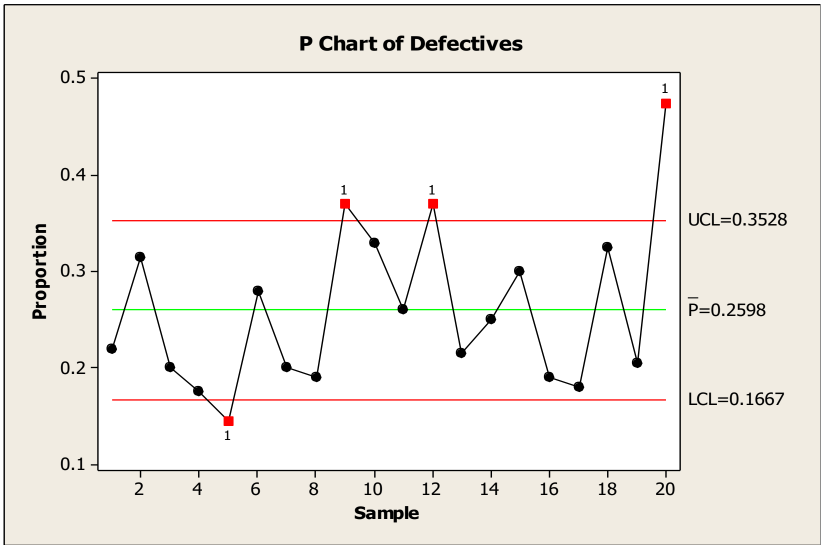


15-55 In a semiconductor manufacturing company, samples of 200 wafers are tested for defectives in the lot. See the number of defectives in 20 such samples in the following table.

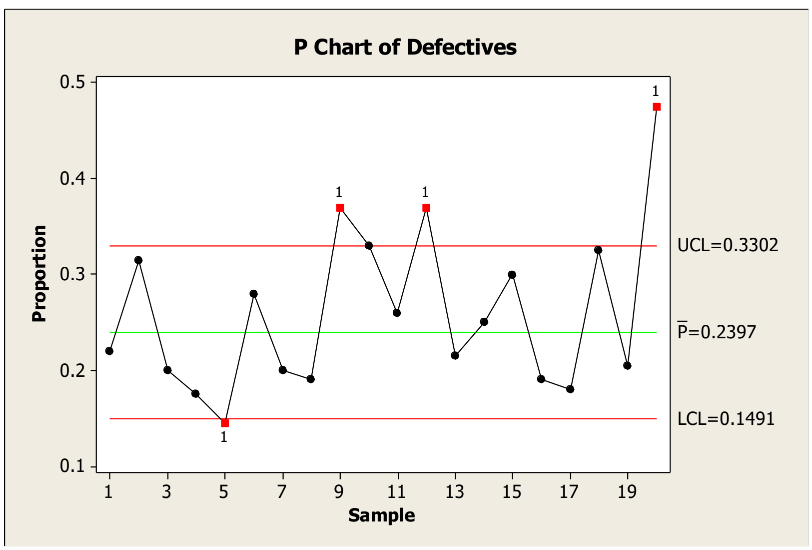
Sample	No. of Defectives	Sample	No. of Defectives
1	44	11	52
2	63	12	74
3	40	13	43
4	35	14	50
5	29	15	60
6	56	16	38
7	40	17	36
8	38	18	65
9	74	19	41
10	66	20	95

(a) Set up a *P* chart for this process. Is the process in statistical control?
(b) Suppose that instead of samples of size 200, we have samples of size 100. Use the data to set up a *P* chart for this process. Revise the control limits if necessary.
(c) Compare the control limits in parts (a) and (b). Explain why they differ.

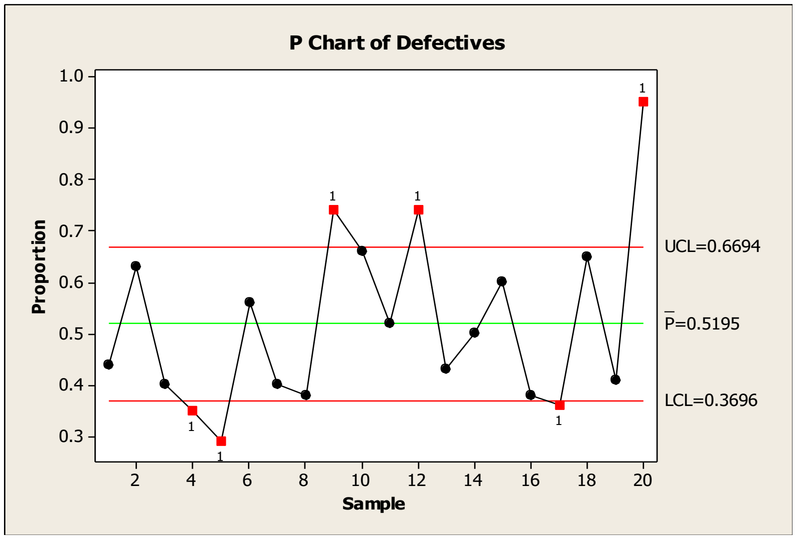
(a) The process is NOT in control. The P chart for subgroups of size 200 follows.



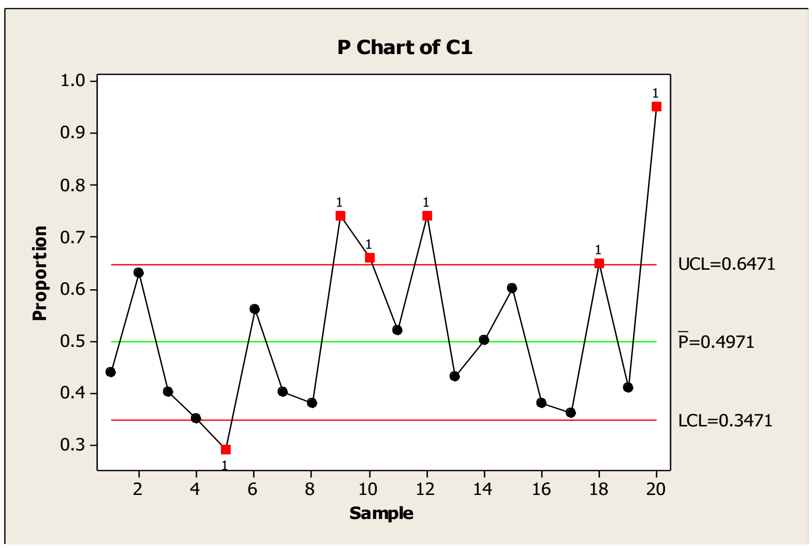
The P chart for subgroups of size 200 follows with points 5, 9, 12, 20 removed from the control limits calculations (but the points are still plotted).



(b) The P chart for subgroups of size 100 follows.



The P chart for subgroups of size 100 follows with points 4, 5, 9, 12, 17, 20 removed from the calculations for the control limits.The control limit is revised one time by omitting the out-of-control points. However, itshows additional out-of-control signals.



(c) The control limits in parts (a) and (b) differ because the subgroup size has changed. The number of defectives is from 200 wafers in part (a) and the same number of defectives is from 100 wafers in part (b). This changes the centerline of the charts and the chart in part (a) also has tighter limits because it uses a larger sample size.

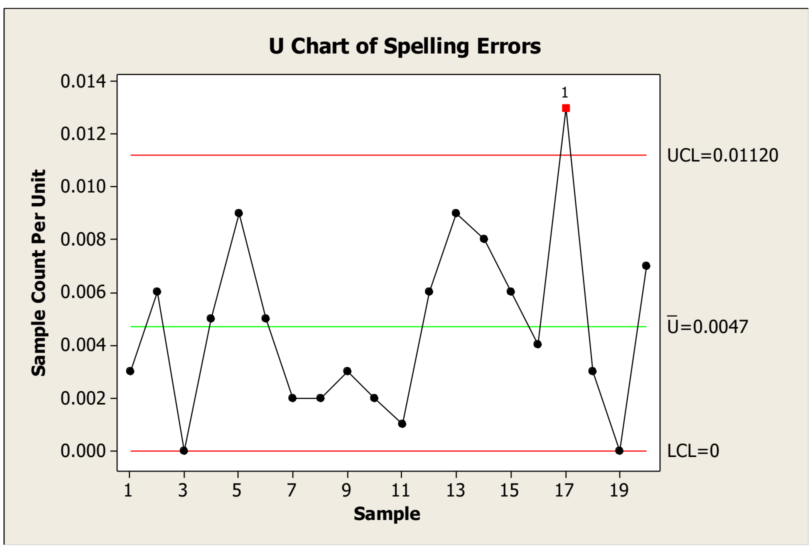
15-56 The following data are the number of spelling errors detected for every 1000 words on a news Web site over 20 weeks.

Week	No. of Spelling Errors	Week	No. of Spelling Errors
1	3	11	1
2	6	12	6
3	0	13	9
4	5	14	8
5	9	15	6
6	5	16	4
7	2	17	13
8	2	18	3
9	3	19	0
10	2	20	7

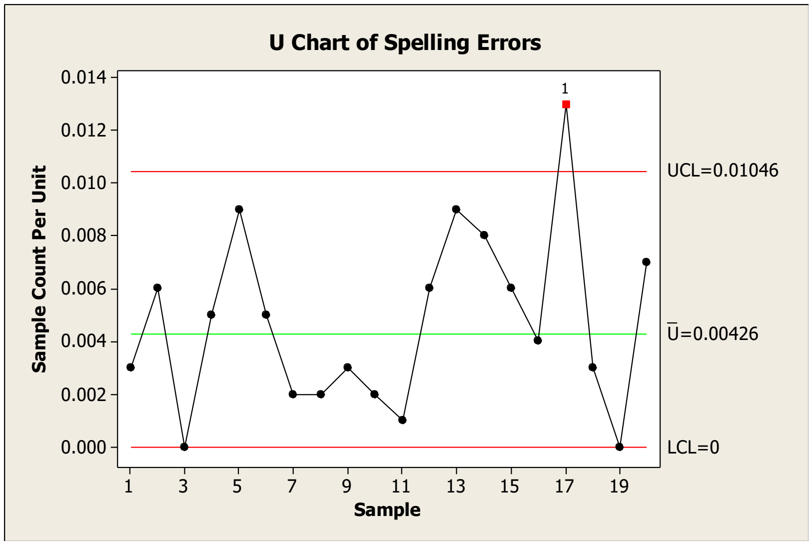
(a) What control chart is most appropriate for these data?
(b) Using all the data, compute trial control limits for the chart in part (a), construct the chart, and plot the data.
(c) Determine whether the process is in statistical control. If not, assume that assignable causes can be found and out-of-control points eliminated. Revise the control limits.

(a) The U chart is appropriate for these data.

(b). The U chart follows.



(c)The process is out-of-control at point 17. The U chart follows with point 17removed from the calculations for the control limits.

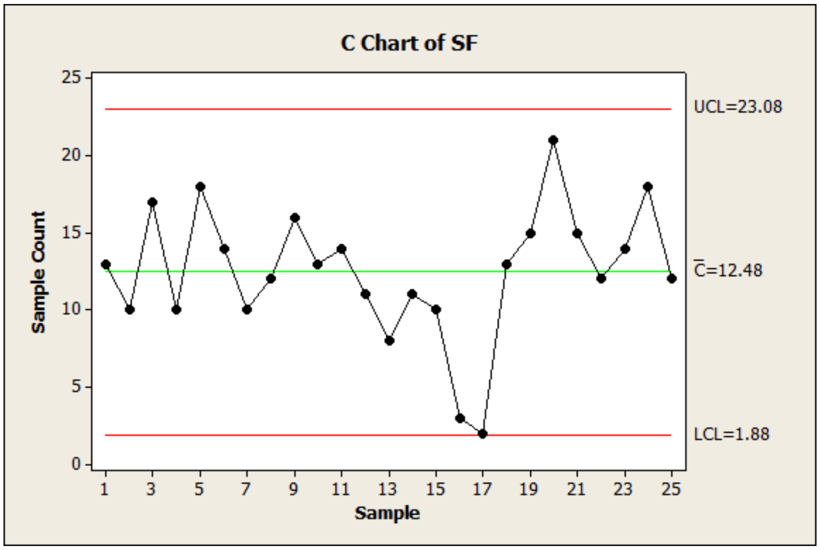


15-57 A article of *Epilepsy Research* ["Statistical Process Control (SPC): A Simple Objective Method for Monitoring Seizure Frequency and Evaluating Effectiveness of Drug Interventions in Refractory Childhood Epilepsy," (2010, Vol 91, pp. 205–213)] used control charts to monitor weekly seizure changes in patients with refractory childhood epilepsy. The following table shows representative data of weekly observations of seizure frequency (SF).

Week	1	2	3	4	5	6	7	8	9	10
SF	13	10	17	10	18	14	10	12	16	13
Week	11	12	13	14	15	16	17	18	19	20
SF	14	11	8	11	10	3	2	13	15	21
Week	21	22	23	24	25					
SF	15	12	14	18	12					

(a) What type of control chart is appropriate for these data? Construct this chart.
(b) Comment on the control of the process.
(c) If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.
(d) In the publication, the weekly SFs were approximated as normally distributed and an individuals chart was constructed. Construct this chart and compare it to the attribute chart you built in part (a).

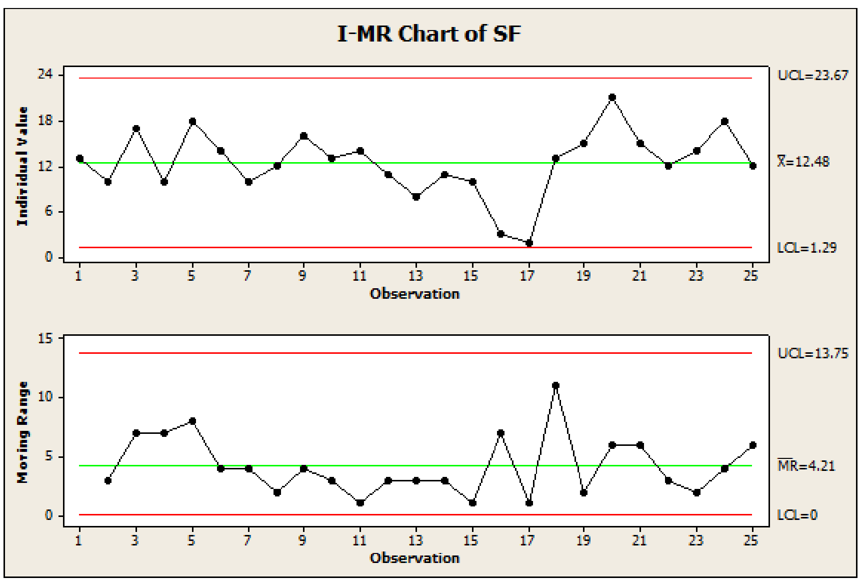
(a) A C-chart is appropriate for these data



(b) The process is in control

(c) No signal is reported and no revision of the control chart is needed

(d) I-chart: UCL=23.67, CL=12.48, LCL=1.29 and both the I-chart and the C-chart indicate that the process is in-control. The normal approximation to the Poisson distribution becomes reasonable for larger Poisson values.



15-58 A article in *Graefe's Archive for Clinical and Experimental Ophthalmology* ["Statistical Process Control Charts for Ophthalmology," (2011, Vol. 249, pp. 1103–1105)] considered the number of cataract surgery cases by month. The data are shown in the following table.

(a) What type of control chart is appropriate for these data? Construct this chart.

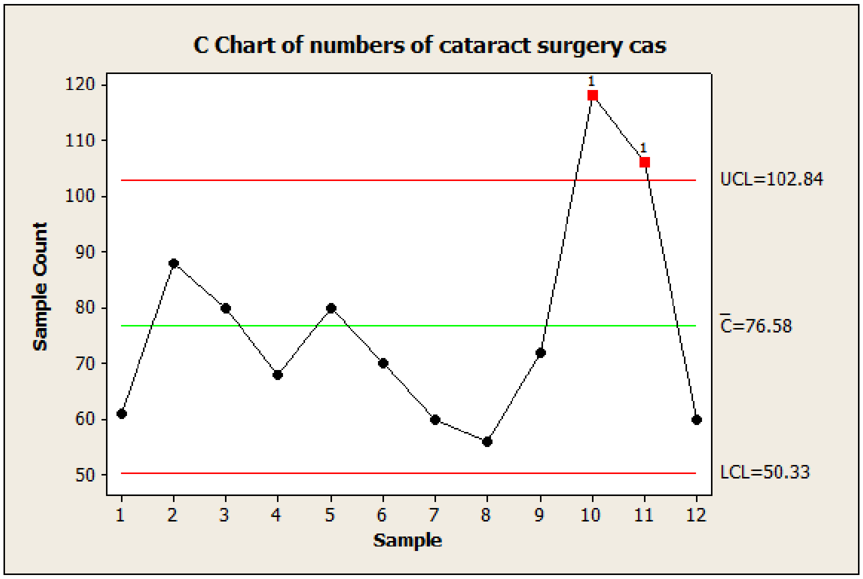
(b) Comment on the control of the process.

(c) If necessary, assume that assignable causes can be found, eliminate suspect points, and revise the control limits.

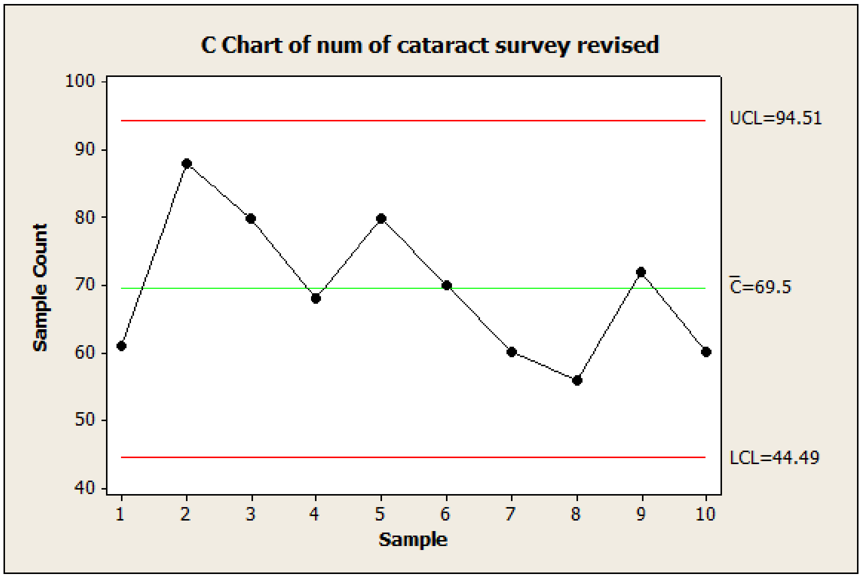
(d) In the publication, the data were approximated as normally distributed and an individuals chart was constructed. Construct this chart and compare it to the attribute chart you built in part (a). Why might an individuals chart be reasonable?

January	February	March	April	May	June	July
61	88	80	68	80	70	60
August	September	October	November	December		
56	72	118	106	60		

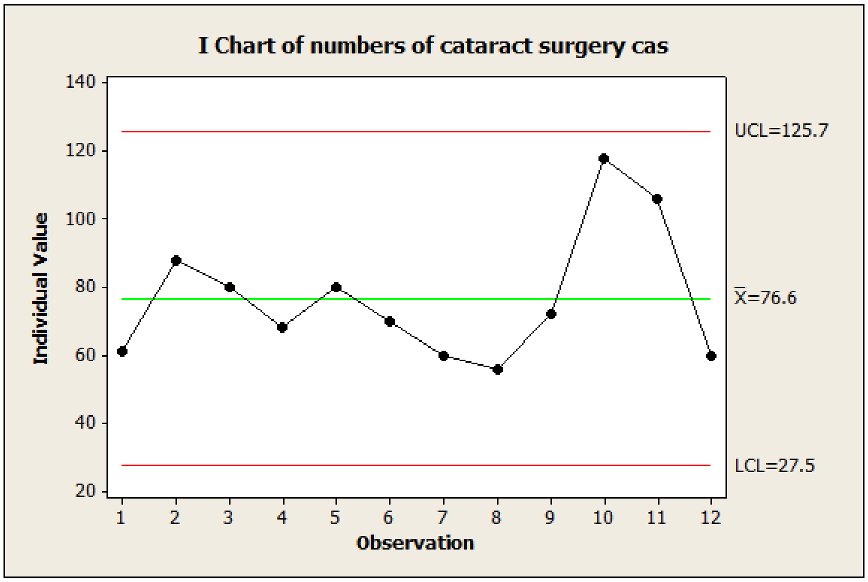
(a) A C-Chart is appropriate for this data.



(b) Two points (points 10 and 11) fall outside 3 standard deviation control limits.
(c) Assume assignable causes can be found, eliminate these suspect points, and revise the control limits.



(d) Individual Chart: no out-of-control points are detected. The normal approximation to the Poisson distribution becomes reasonable for larger Poisson values.



Because the data is approximately normal distributed and with weak autocorrelation, individual chart is also a reasonable choice.

Section 15-7

15-59 An X chart uses samples of size 1. The center line is at 100, and the upper and lower 3-sigma limits are at 112 and 88, respectively.

(a) What is the process σ?
(b) Suppose that the process mean shifts to 96. Find the probability that this shift is detected on the next sample.
(c) Find the ARL to detect the shift in part (b).

(a) (112 – 100/3=4
(b)

$$P(88 < X < 112) = P\left(\frac{88-96}{4} < \frac{X-\mu}{\sigma_X} < \frac{112-96}{4}\right) = P\left(-2 < Z < 4\right)$$
$$= P(Z < 4) - P(Z < -2) = 0.9772$$

The probability of detecting is 1-0.9772=0.0228.

(c) 1/0.0228=43.8596 ARL to detect the shift is about 44.

15-60 An $\bar{X}$ chart uses samples of size 4. The center line is at 100, and the upper and lower 3-sigma control limits are at 106 and 94, respectively.

(a) What is the process σ?
(b) Suppose that the process mean shifts to 96. Find the probability that this shift is detected on the next sample.
(c) Find the ARL to detect the shift in part (b).

(a) $\mu + 3\frac{\sigma}{\sqrt{n}} = UCL$

$$100 + 3\frac{\sigma}{\sqrt{4}} = 106$$

$$\sigma = \frac{2}{3}(106 - 100) = 4$$

(b) $\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{4}{2} = 2, \qquad \mu = 96$

$$P(94 < X < 106) = P\left(\frac{94-96}{2} < \frac{\bar{X}-\mu}{\sigma_{\bar{x}}} < \frac{106-96}{2}\right) = P\left(-1 < Z < 5\right)$$
$$= P(Z < 5) - P(Z < -1) = 1 - 0.1587 = 0.8413$$

The probability that this shift will be detected on the next sample is p = 1–0.8413 = 0.1587.

(c) $ARL = \frac{1}{p} = \frac{1}{0.1587} = 6.301$

15-61 Consider the $\bar{X}$ control chart in Fig. 15-3. Suppose that the mean shifts to 74.010 millimeters.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\bar{x} = 74.01$ $\sigma_{\bar{x}} = 0.0045$ $\mu = 74.01$

$$P(73.9865 < \overline{X} < 74.0135)$$
$$= P\left(\frac{73.9865 - 74.01}{0.0045} < \frac{X-\mu}{\hat{\sigma}_{\bar{x}}} < \frac{74.0135 - 74.01}{0.0045}\right)$$
$$= P(-5.22 < Z < 0.78) = P(Z < 0.78) - P(Z < -5.22)$$
$$= P(Z < 0.78) - [1 - P(Z < 5.22)] = 0.7823 - (1-1) = 0.7823$$

The probability that this shift will be detected on the next sample is p = 1–0.7823 = 0.2177.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.2177} = 4.6$

15-62 Consider an $\bar{X}$ control chart with $\bar{r}$ = 0.344, UCL = 14.708, LCL = 14.312, and n = 5. Suppose that the mean shifts to 14.6.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\hat{\sigma} = \dfrac{\overline{R}}{d_2} = \dfrac{0.344}{2.326} = 0.148$ $\hat{\sigma}_{\bar{x}} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{0.148}{\sqrt{5}} = 0.066$, $\mu = 14.6$

$$P(14.312 < X < 14.708) = P\left(\frac{14.312 - 14.6}{0.066} < \frac{X-\mu}{\sigma_x} < \frac{14.708 - 14.6}{0.066}\right)$$
$$= P(-4.36 < Z < 1.64) = P(Z < 1.64) - P(Z < -4.36)$$
$$= 0.94950 - (0) = 0.94950$$

The probability that this shift will be detected on the next sample is p = 1–0.94950 = 0.0505.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.0505} = 19.8$

15-63 Consider an $\bar{X}$ control chart with $\bar{r}$ = 34.286, UCL = 242.780, LCL = 203.220, and n = 5. Suppose that the mean shifts to 210.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\hat{\sigma} = \dfrac{\overline{R}}{d_2} = 14.74$ $\hat{\sigma}_{\bar{x}} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{14.74}{\sqrt{5}} = 6.592$, $\mu = 210$

$$P(203.22 < \overline{X} < 242.78) = P\left(\frac{203.22 - 210}{6.592} < \frac{\overline{X}-\mu}{\sigma_{\bar{x}}} < \frac{242.78 - 210}{6.592}\right)$$
$$= P(-1.03 < Z < 4.97) = P(Z < 4.97) - P(Z < -1.03)$$
$$= 1 - 0.1515 = 0.8485$$

The probability that this shift will be detected on the next sample is p = 1–.8485= 0.1515.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.1515} = 6.6$

15-64 Consider an $\bar{X}$ control chart with $\hat{\sigma}$ = 1. 40, UCL = 21.71, LCL = 18.29, and n = 6. Suppose that the mean shifts to 17.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\hat{\sigma} = \dfrac{\overline{R}}{d_2} = 1.4$ $\hat{\sigma}_{\overline{x}} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{1.4}{\sqrt{6}} = 0.5715$, μ = 17

$$P(18.29 < X < 21.71) = P\left(\frac{18.29-17}{0.5715} < \frac{\overline{X}-\mu}{\sigma_{\overline{x}}} < \frac{21.71-17}{0.5715}\right)$$
$$= P(2.26 < Z < 8.24) = P(Z < 8.24) - P(Z < 2.26)$$
$$= 1 - 0.9881 = 0.0119$$

The probability that this shift will be detected on the next sample is p = 1–0.0119 = 0.9881.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.9881} = 1.012$

15-65 Consider an $\overline{X}$ control chart with $\hat{\sigma}$ = 2.466, UCL = 37.404, LCL = 30.780, and n = 5. Suppose that the mean shifts to 36.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\hat{\sigma}_{\overline{x}} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{2.4664}{\sqrt{5}} = 1.103$, μ = 36

$$P(30.78 < \overline{X} < 37.404) = P\left(\frac{30.78-36}{1.103} < \frac{\overline{X}-\mu}{\hat{\sigma}_{\overline{x}}} < \frac{37.404-36}{1.103}\right)$$
$$= P(-4.73 < Z < 1.27) = P(Z < 1.27) - P(Z < -4.73)$$
$$= 0.8980 - 0 = 0.8980$$

The probability that this shift will be detected on the next sample is p = 1–0.8980 = 0.1020.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.102} = 9.8$

15-66 Consider an $\overline{X}$ control chart with $\bar{r}$ = 2.25, UCL =17.40, LCL = 12.79, and n = 3. Suppose that the mean shifts to 13.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\hat{\sigma} = \dfrac{\overline{R}}{d_2} = \dfrac{2.25}{1.693} = 1.329$ $\hat{\sigma}_{\overline{x}} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{1.329}{\sqrt{3}} = 0.767$, μ = 13

$$P(12.79 < \overline{X} < 17.4) = P\left(\frac{12.79-13}{0.767} < \frac{\overline{X}-\mu}{\sigma_{\overline{x}}} < \frac{17.4-13}{0.767}\right)$$
$$= P(-0.27 < Z < 5.74) = P(Z < 5.74) - P(Z < -0.27)$$
$$= 1 - 0.3936 = 0.6064$$

The probability that this shift will be detected on the next sample is p = 1–06064 = 0.3936.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.3936} = 2.54$

15-67 Consider an $\overline{X}$ control chart with $\bar{r}$ = 0.00924, *UCL* = 0.0635, *LCL* = 0.0624, and *n* = 5. Suppose that the mean shifts to 0.0625.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\hat{\sigma} = \dfrac{\overline{R}}{d_2} = \dfrac{0.000924}{2.326} = 0.000397$ $\hat{\sigma}_{\bar{x}} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{0.000397}{\sqrt{5}} = 0.000178$, μ = 0.0625

$$P(0.0624 < \overline{X} < 0.0635) = P\left(\frac{0.0624-0.0625}{0.000178} < \frac{\overline{X}-\mu}{\sigma_{\bar{x}}} < \frac{0.0635-0.0625}{0.000178}\right)$$
$$= P(-0.56 < Z < 5.62) = P(Z < 5.62) - P(Z < -0.56)$$
$$= 1 - 0.2877 = 0.7123$$

The probability that this shift will be detected on the next sample is *p* = 1 – 0.7123 = 0.2877.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.2877} = 3.48$

15-68 Consider the revised $\overline{X}$ control chart in Exercise 15-8 with $\hat{\sigma}$ = 0.669, *UCL* = 7.443, *LCL* = 5.125, and *n* = 3. Suppose that the mean shifts to 5.5.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL after the shift?

(a) $\hat{\sigma} = \dfrac{\overline{R}}{d_2} = 0.669$ $\hat{\sigma}_{\bar{x}} = \dfrac{\hat{\sigma}}{\sqrt{n}} = \dfrac{0.669}{\sqrt{3}} = 0.386$, μ = 5.5

$$P(5.125 < \overline{X} < 7.443 \mid \mu = 5.5) = P\left(\frac{5.125-5.5}{0386} < \frac{\overline{X}-\mu}{\sigma_{\bar{x}}} < \frac{7.443-5.5}{0.386}\right)$$
$$= P(-0.97 < Z < 5.03) = P(Z < 5.03) - P(Z < -0.97)$$
$$= 1 - 0.16603 = 0.83397$$

The probability that this shift will be detected on the next sample is *p* = 1–0.83397 = 0.16603.

(b) $ARL = \dfrac{1}{p} = \dfrac{1}{0.16603} = 6.02$

15-69 An $\overline{X}$ chart uses a sample of size 3. The center line is at 200, and the upper and lower 3-sigma control limits are at 212 and 188, respectively.

(a) Estimate the process σ.
(b) Suppose that the process mean shifts to 195. Determine the probability that this shift is detected on the next sample.
(c) Find the ARL to detect the shift in part (b).

(a) The difference $UCL - LCL = 6\hat{\sigma}_{\overline{X}} = 212 - 188 = 24$
Therefore, $\hat{\sigma}_{\overline{X}} = \frac{24}{6} = 4$ and $\hat{\sigma}_{\overline{X}} = \frac{\hat{\sigma}}{\sqrt{3}}$. Therefore, $\hat{\sigma} = 4\sqrt{3} = 6.928$

(b) $P(188 < \overline{X}$< 212 | μ = 195) = P[(188 – 195)/4< Z < (212 – 195)/4] = P(-1.75< Z <4.25) = 0.9599.
Therefore, the probability the shift is detected = 1 – 0.9902 = 0.0401 ≈ 0.04

(c) ARL = 1/*p*, where *p* is the probability a point exceeds a control limit. From part (b) *p* = 0.0401. Therefore ARL = 1/0.0401 = 24.94.

15-70 Consider an $\overline{X}$ control chart with *UCL* = 24.802, *LCL* = 23.792, and *n* = 3. Suppose that the mean shifts to 24.2.

(a) What is the probability that this shift is detected on the next sample?

(b) What is the ARL after the shift?

(a) The difference $UCL - LCL = 6\hat{\sigma}_{\overline{X}} = 24.802 - 23.792 = 1.01$

Therefore, $\hat{\sigma}_{\overline{X}} = \frac{1.01}{6} = 0.1683$ and $\hat{\sigma} = \frac{1.01\sqrt{3}}{6} = 0.2916$

$$P(23.792 < \overline{X} < 24.802 \mid \mu = 24.2) = P\left(\frac{23.792 - 24.2}{0.1683} < \frac{\overline{X} - \mu}{\sigma_{\bar{x}}} < \frac{24.802 - 24.2}{0.1683}\right)$$
$$= P(-2.4242 < Z < 3.5769) = P(Z < 3.5769) - P(Z < -2.4242)$$
$$= 0.9998 - 0.0077 = 0.9921$$

Therefore the probability the shift is detected = 1 – 0.9921 = 0.0079 ≈ 0.008.

(c) ARL = 1/*p*, where *p* is the probability a point exceeds a control limit. From part (b),*p* = 0.0079. Therefore ARL = 1/0.0079 = 126.58

15-71 Consider a *P*-chart with subgroup size *n* = 50 and center line at 0.12.

(a) Calculate the *LCL* and *UCL*.
(b) Suppose that the true proportion defective changes from 0.12 to 0.18. What is the ARL after the shift? Assume that the sample proportions are approximately normally distributed.
(c) Rework part (a) and (b) with *n* = 100 and comment on the difference in ARL. Does the increased sample size change the ARL substantially?

(a) $= p + 3\sqrt{\frac{p(1-p)}{n}} = 0.258$, $UCL = p - 3\sqrt{\frac{p(1-p)}{n}} = -0.018$, set to zero

(b) Assume p is normally distributed $p \sim N(0.18, \sigma^2)$ and $\hat{\sigma} = \sqrt{\frac{p(1-p)}{n}} = 0.46$

$P(\hat{p} < LCL) < 0.0001$, $P(\hat{p} > UCL) = 0.045$, $ARL = \frac{1}{0.045} = 22.1$

(c) When$n = 100$, $= p + 3\sqrt{\frac{p(1-p)}{n}} = 0.217$, $UCL = p - 3\sqrt{\frac{p(1-p)}{n}} = 0.023$

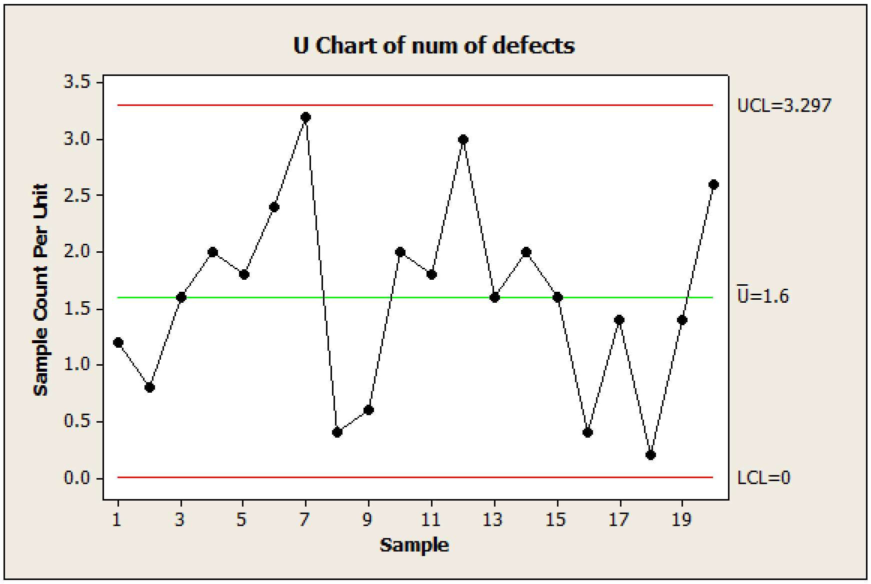
Assume p is normally distributed $p \sim N(0.18, \sigma^2)$ and $\hat{\sigma} = \sqrt{\frac{p(1-p)}{100}} = 0.325$, $P(\hat{p} < LCL) < 0.0001$, $P(\hat{p} > UCL) = P(\hat{p} > \frac{0.217-0.18}{0.325}) = 0.124$, $ARL = \frac{1}{0.124} = 8.04$

The change in sample size changes the ARL substantially.

15-72 Consider the *U* chart for printed circuit boards in Example 15-5. The center line = 1.6, *UCL* = 3 3 . , and *n* = 5.

(a) Calculate the *LCL* and *UCL*.
(b) Suppose that the true mean defects per unit shifts from 1.6 to 2.4. What is the ARL after the shift? Assume that the average defects per unit are approximately normally distributed.
(c) Rework part (b) if the true mean defects per unit shifts from 1.6 to 2.0 and comment on the difference in ARL.

(a)



$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} = 1.6 + 3\sqrt{\frac{1.6}{5}} = 3.3$$

$$LCL = \max\left(0, \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} = -0.09\right) = 0$$

(b) The standard deviation of U is $\sqrt{\frac{\mu}{n}} = \sqrt{\frac{2.4}{5}} = 0.692$ when the mean of U is 2.4. Therefore, $P(U < LCL) = 0$ and $P(U > UCL) \approx P(Z > \frac{3.3-2.4}{0.692}) = 0.098$, $ARL = \frac{1}{0.098} = 10.23$

(c) The standard deviation of U is $\sqrt{\frac{\bar{u}}{n}} = \sqrt{\frac{2.0}{5}} = 0.632$, when the mean of U is 2.0. Therefore, $P(U < LCL) = 0$, $P(U > UCL) \approx P(Z > \frac{3.3-2.0}{0.632}) = 0.202$, $ARL = \frac{1}{0.202} = 49.65$

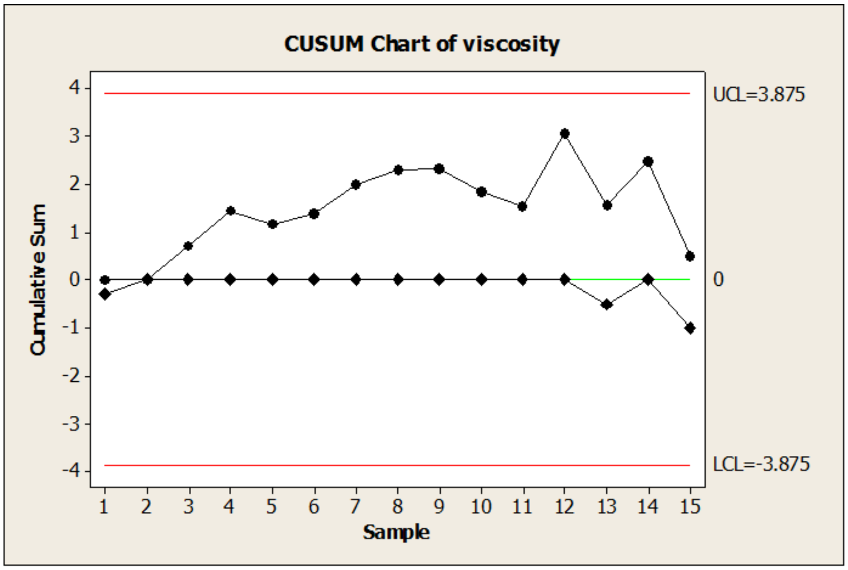
Section 15-8

15-73 The following data were considered in *Quality Engineering* ["Parabolic Control Limits for The Exponentially Weighted Moving Average Control Charts in Quality Engineering" (1992, Vol. 4(4)]. In a chemical plant, the data for one of the quality characteristics (viscosity) were obtained for each 12-hour batch's at the batch completion. The results of 15 consecutive measurements are shown in the following table.

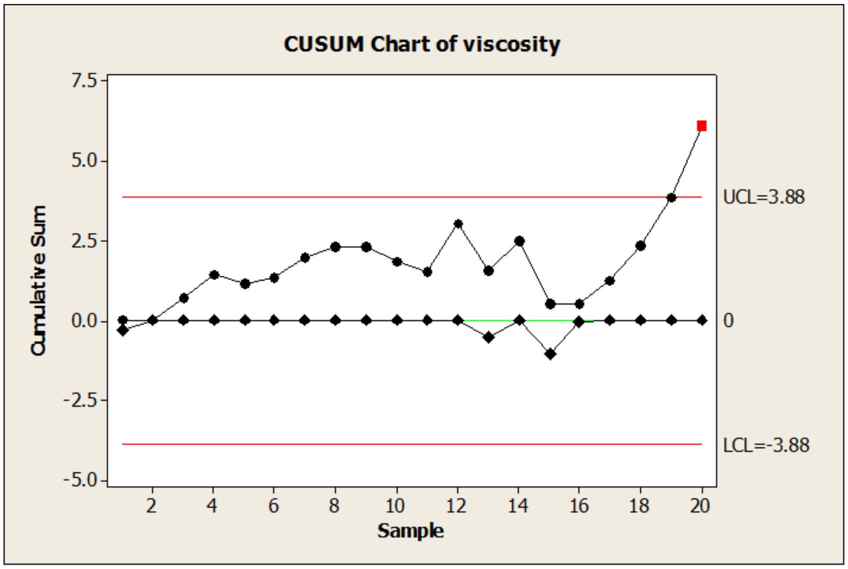
Batch	Viscosity	Batch	Viscosity
1	13.3	9	14.6
2	14.5	10	14.1
3	15.3	11	14.3
4	15.3	12	16.1
5	14.3	13	13.1
6	14.8	14	15.5
7	15.2	15	12.6
8	14.9		

(a) Set up a CUSUM control chart for this process. Assume that the desired process target is 14.1. Does the process appear to be in control?
(b) Suppose that the next five observations are 14.6, 15.3, 15.7, 16.1, and 16.8. Apply the CUSUM in part (a) to these new observations. Is there any evidence that the process has shifted out of control?

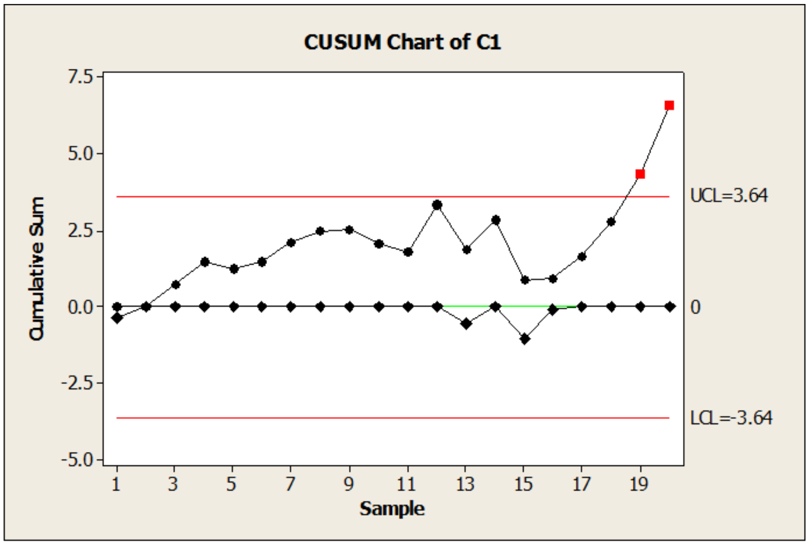
(a) Yes, this process is in-control.CUSUM chart with h = 4 and k = 0.5 is shown.



(b) Yes, this process has shifted out-of-control. For the CUSUM estimated from all the data (with k = 0.5 and h = 4) observation 20 exceeds the upper limit.



For a CUSUM with standard deviation estimated from the moving range of the first 15 observation, the moving range is 1.026 and the standard deviation estimate is 0.9096. If this standard deviation is used with a target of 14.1, the following CUSUM is obtained and a signal occurs at an earlier observation.

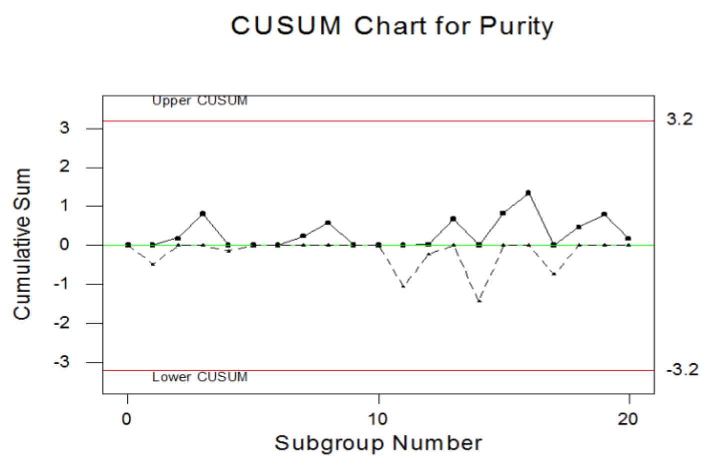


15-74 The purity of a chemical product is measured every two hours. The results of 20 consecutive measurements are as follows:

Sample	Purity	Sample	Purity
1	89.11	11	88.55
2	90.59	12	90.43
3	91.03	13	91.04
4	89.46	14	88.17
5	89.78	15	91.23
6	90.05	16	90.92
7	90.63	17	88.86
8	90.75	18	90.87
9	89.65	19	90.73
10	90.15	20	89.78

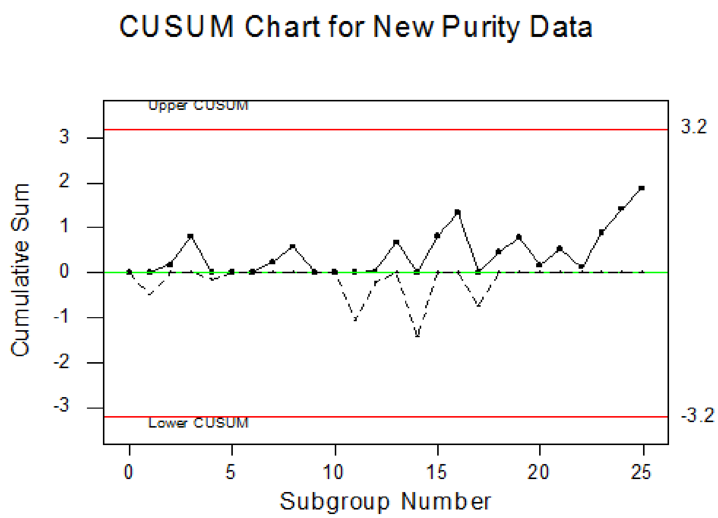
(a) Set up a CUSUM control chart for this process. Use $\sigma = 0.8$ in setting up the procedure, and assume that the desired process target is 90. Does the process appear to be in control?
(b) Suppose that the next five observations are 90.75, 90.00, 91.15, 90.95, and 90.86. Apply the CUSUM in part (a) to these new observations. Is there any evidence that the process has shifted out of control?

(a) CUSUM Control chart with k=0.5 and h=4



The CUSUM control chart for purity does not indicate an out-of-control situation. The S_H values do not plot beyond the values of $-H$ and H.

(b) CUSUM Control chart with k=0.5 and h=4



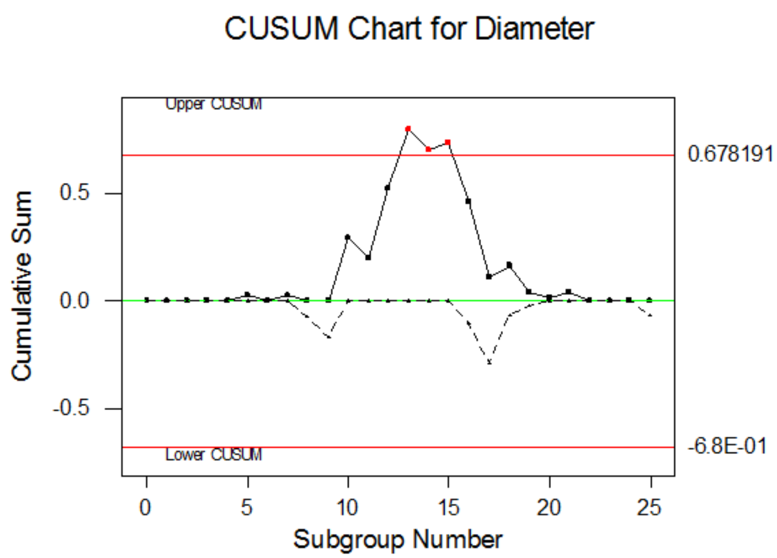
The process appears to be moving out of statistical control.

15-75 An automatic senson measures the diameter of holes in consecutive order. The results of measuring 25 holes follow.

Sample	Diameter	Sample	Diameter
1	9.94	14	9.99
2	9.93	15	10.12
3	10.09	16	9.81
4	9.98	17	9.73
5	10.11	18	10.14
6	9.99	19	9.96
7	10.11	20	10.06
8	9.84	21	10.11
9	9.82	22	9.95
10	10.38	23	9.92
11	9.99	24	10.09
12	10.41	25	9.85
13	10.36		

(a) Estimate the process standard deviation.
(b) Set up a CUSUM control procedure, assuming that the target diameter is 10.0 millimeters. Does the process appear to be operating in a state of statistical control at the desired target level?

(a) $\hat{\sigma}$ = s = 0.1736, estimate from the moving range = 0.1695
(b) CUSUM Control chart with k=0.5 and h=4



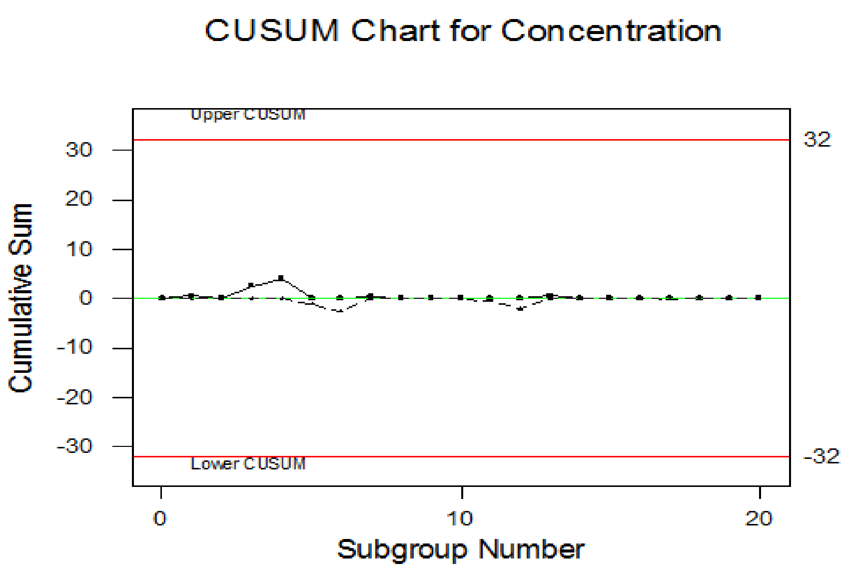
The process appears to be out of control at the specified target level.

15-76 The concentration of a chemical product is measured by taking four samples from each batch of material. The average concentration of these measurements for the last 20 batches is shown in the following table:

Batch	Concentration	Batch	Concentration
1	104.5	11	95.4
2	99.9	12	94.5
3	106.7	13	104.5
4	105.2	14	99.7
5	94.8	15	97.7
6	94.6	16	97
7	104.4	17	95.8
8	99.4	18	97.4
9	100.3	19	99
10	100.3	20	102.6

(a) Suppose that the process standard deviation is σ = 8 and that the target value of concentration for this process is 100. Design a CUSUM scheme for the process. Does the process appear to be in control at the target?
(b) How many batches would you expect to be produced with off-target concentration before it would be detected by the CUSUM control chart if the concentration shifted to 104? Use Table 15-10.

(a) CUSUM Control chart with k=0.5 and h=4



The process appears to be in statistical control.

(b) With the target = 100,a shift to 104 is a shift of 104 – 100 = 4 = 0.5σ. From Table 16-9 with h = 4 and a shift of 0.5, ARL = 26.6

15-77 Consider a CUSUM with h = 5 and k = 1 / 2. Samples are taken every two hours from the process. The target value for the process is $\mu_0 = 50$ and $\sigma = 2$. Use Table 15-10.

(a) If the sample size is n = 1, how many samples would be required to detect a shift in the process mean to $\mu = 51$ on average?

(b) How does increasing the sample size to n = 4 affect the average run length to detect the shift to $\mu = 51$ that you determined in part (a)?

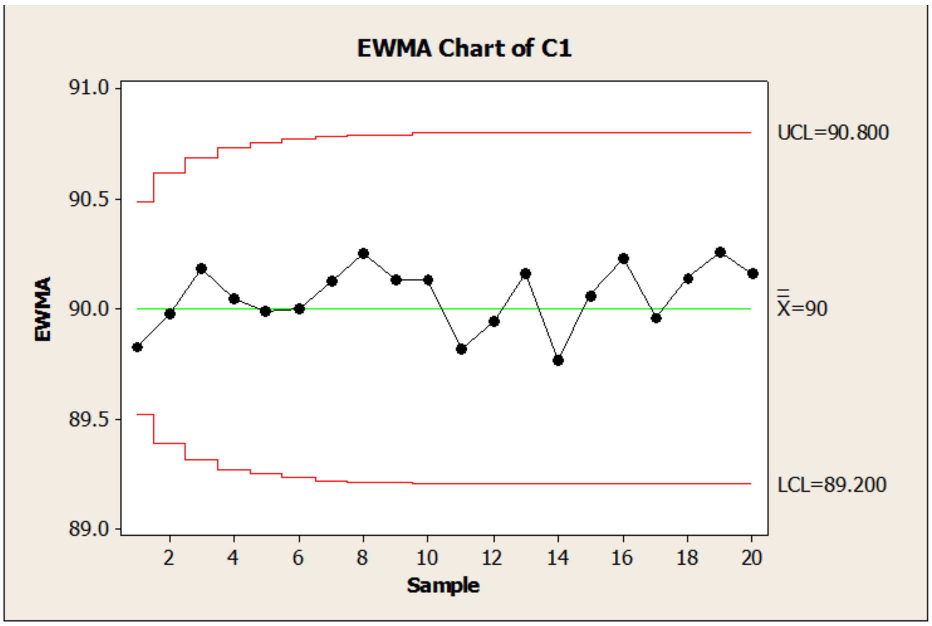
(a) A shift to 51 is a shift of $\frac{\mu-\mu_0}{\sigma}=\frac{51-50}{2}$ = 0.5 standard deviations. From Table 16-9, ARL = 38.0

(b) If n = 4, the shift to 51 is a shift of $\frac{\mu-\mu_0}{\sigma/\sqrt{n}}=\frac{51-50}{2/\sqrt{4}}$ = 1 standard deviation. From Table 16-9, ARL = 10.4

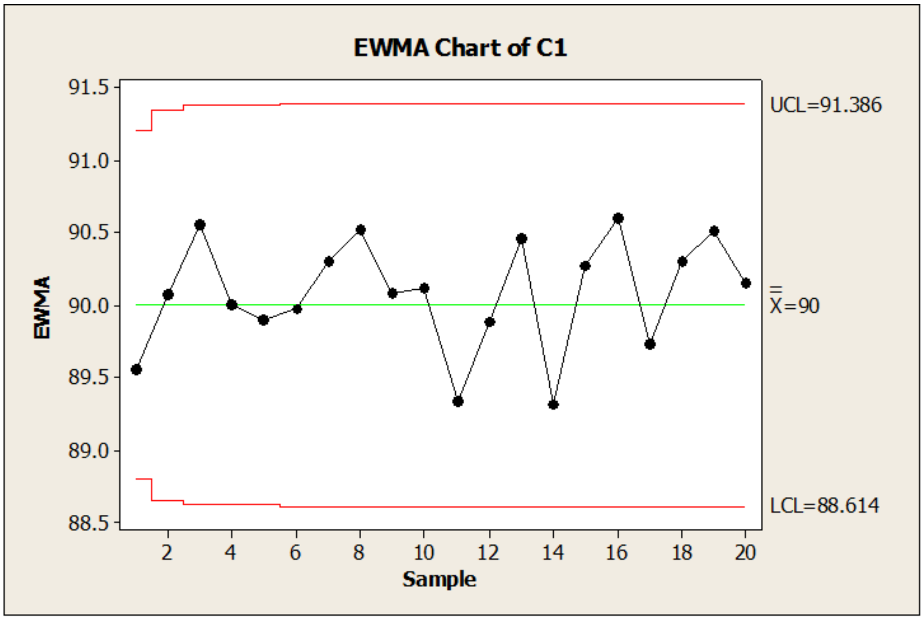
15-78 Consider the purity data in Exercise 15-74. Use $\sigma = 0.8$ and assume that the desired process target is 90.

(a) Construct an EWMA control chart with $\lambda = 0.2$. Does the process appear to be in control?
(b) Construct an EWMA control chart with $\lambda = 0.5$. Compare your results to part (a).
(c) Suppose that the next five observations are 90.75, 90.00, 91.15, 90.95, and 90.86. Apply the EWMAs in part (a) and (b) to these new observations. Is there any evidence that the process has shifted out of control?

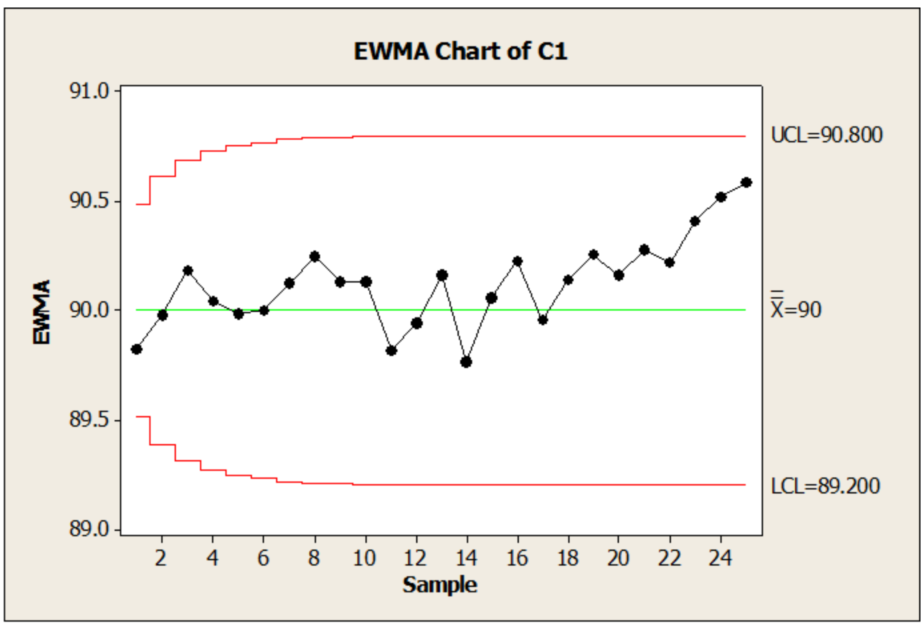
(a) The process appears to be in control.



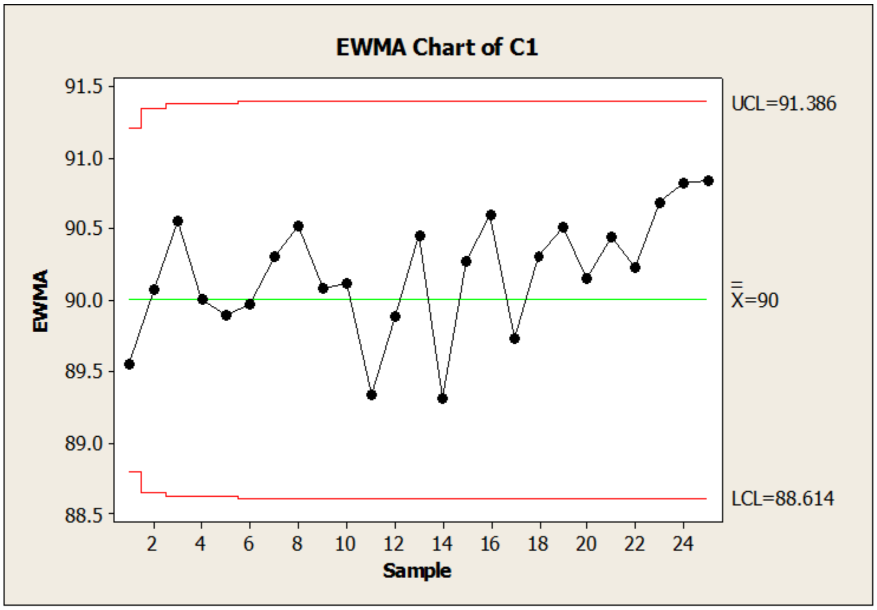
(b) The process appears to be in control.



(c) For part (a), there is no evidence that the process has shifted out of control.



For part b), there is no evidence that the process has shifted out of control.

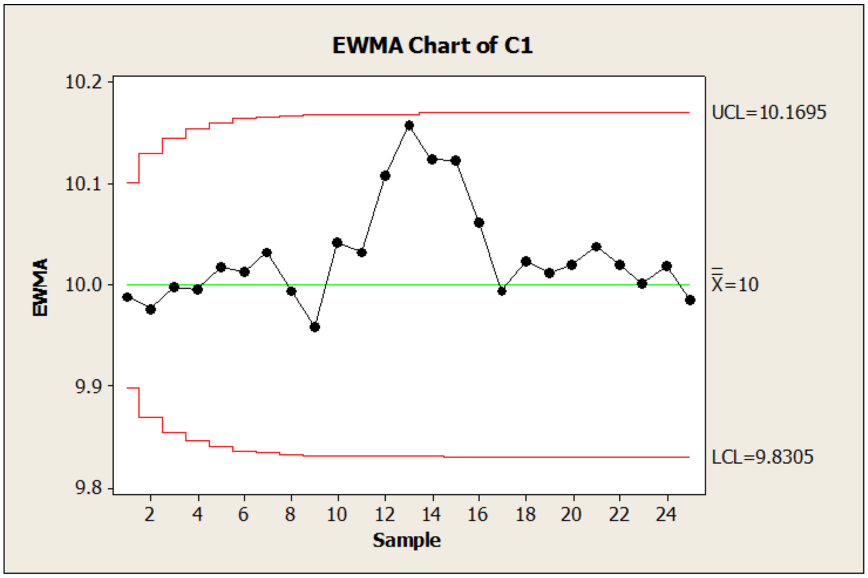


15-79 Consider the diameter data in Exercise 15-75. Assume that the desired process target is 10.0 millimeters.

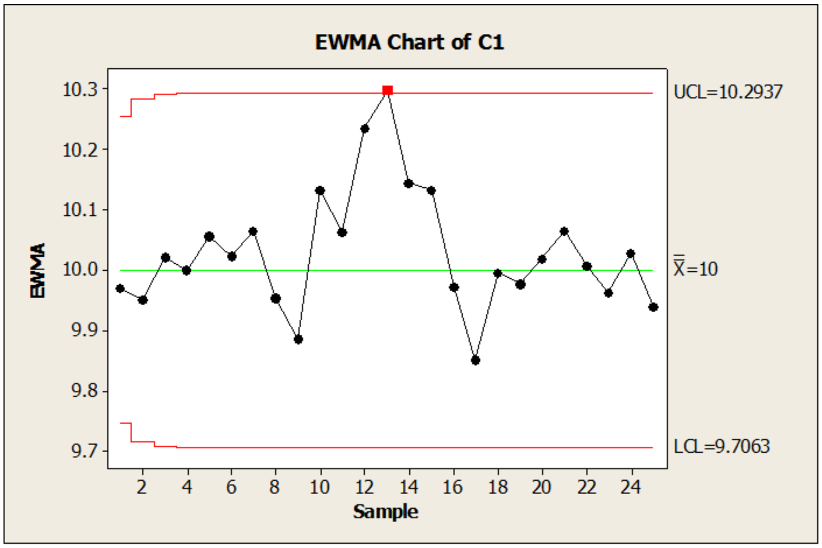
(a) Estimate the process standard deviation.
(b) Construct an EWMA control chart with $\lambda=0.2$. Does the process appear to be in control?
(c) Construct an EWMA control chart with $\lambda=0.5$. Compare your results to part (a).

(a) Estimated standard deviation from the moving range = 0.1695.

(b) The process appears to be in control.



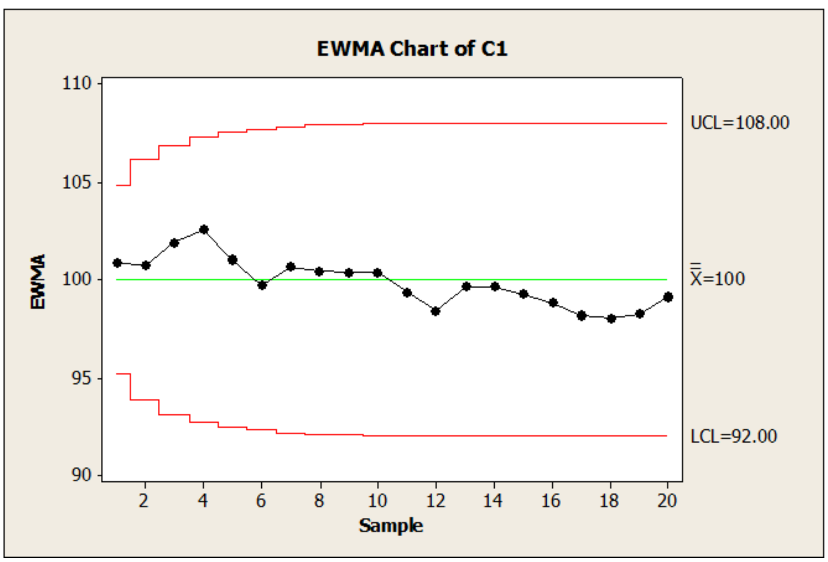
(c) The process appears to be out of control at the observation 13.



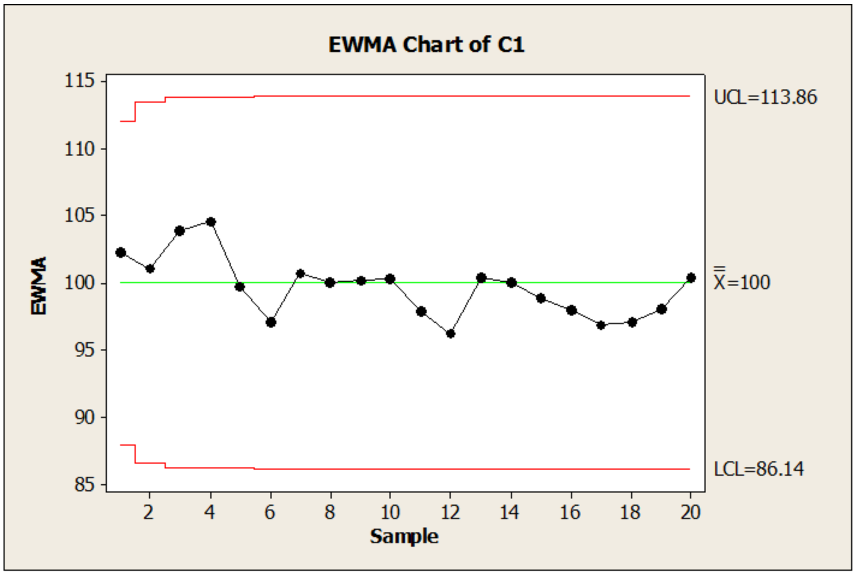
15-80 Consider the concentration data in Exercise 15-76. Use $\sigma = 8$ and assume that the desired process target is 100.

(a) Construct an EWMA control chart with $\lambda = 0.2$. Does the process appear to be in control?
(b) Construct an EWMA control chart with $\lambda = 0.5$. Compare your results to part (a).
(c) If the concentration shifted to 104, would you prefer the chart in part (a) or (b)? Explain.

(a) The process appears to be in control.



(b) The process appears to be in control.



(c) Because the shift is 0.5σ, a smaller λ is preferred to detect the shift quickly. Therefore, the chart in part (a) is preferred.

15-81 Consider an EMWA control chart. The target value for the process is $\mu_0 = 50$ and $\sigma = 2$. Use Table 15-11.

(a) If the sample size is n = 1, would you prefer an EWMA chart with λ=0.1 and L = 2.81 or λ=0.5 and L = 3.07 to detect a shift in the process mean to $\mu = 52$ on average? Why?
(b) If the sample size is increased to n = 4, which chart in part (a) do you prefer? Why?
(c) If an EWMA chart with λ=0.1 and L = 2.81 is used, what sample size is needed to detect a shift to $\mu = 52$ in approximately three samples on average?

(a) The shift of the mean is 1σ. So we prefer $\lambda = 0.1$ and L=2.81 because this setting has thesmaller ARL = 10.3.
(b) The shift of the mean is $2\sigma_{\overline{X}}$. So we prefer $\lambda = 0.5$ and L=3.07 because this setting has the smaller ARL = 3.63

(c) The shift of the mean is $3\sigma_{\overline{X}}$. Solving $2/\left(\dfrac{2}{\sqrt{n}}\right) = 3$ for n gives us the required sample size of 9.

15-82 A process has a target of $\mu_0 = 100$ and a standard deviation of $\sigma = 2$. Samples of size n = 1 are taken every two hours. Use Table 15-10.

(a) Suppose that the process mean shifts to $\mu = 102$. How many hours of production occur before the process shift is detected by a CUSUM with h = 5 and k = 1 / 2?
(b) It is important to detect the shift defined in part (a) more quickly. A proposal to reduce the sampling frequency to 0.5 hour is made. How does this affect the CUSUM control procedure? How much more quickly is the shift detected?
(c) Suppose that the 0.5 hour sampling interval in part (b) is adopted. How often do false alarms occur's with this new sampling interval? How often did they occur with the old interval of two hours?
(d) A proposal is made to increase the sample size to n = 4 and retain the two-hour sampling interval. How does this suggestion compare in terms of average detection time to the suggestion of decreasing the sampling interval to 0.5 hour?

(a) With a target = 100 and a shift to 102 results in a shift of $\dfrac{102-100}{4}$ = 0.5 standard deviations.
From Table 16-9, ARL = 38. The hours of production are 2(38) = 76.

(b) The ARL = 38. However, the time to obtain 38 samples is now 0.5(38) = 19.

(c) From Table 16-9, the ARL when there is no shift is 465. Consequently, the time between false alarms is 0.5(465) = 232.5 hours. Under the old interval, false alarms occurred every 930 hours.

(d) If the process shifts to 102, the shift is $\frac{\mu-\mu_0}{\sigma/\sqrt{n}} = \frac{102-100}{4/\sqrt{4}} = 1$ standard deviation. From Table 16-9, theARL for this shift is 10.4. Therefore, the time to detect the shift is 2(10.4) = 20.8 hours. Although this time is slightly longer than the result in part (b), the time between false alarms is 2(465) = 930 hours, which isbetter than the result in part (c).

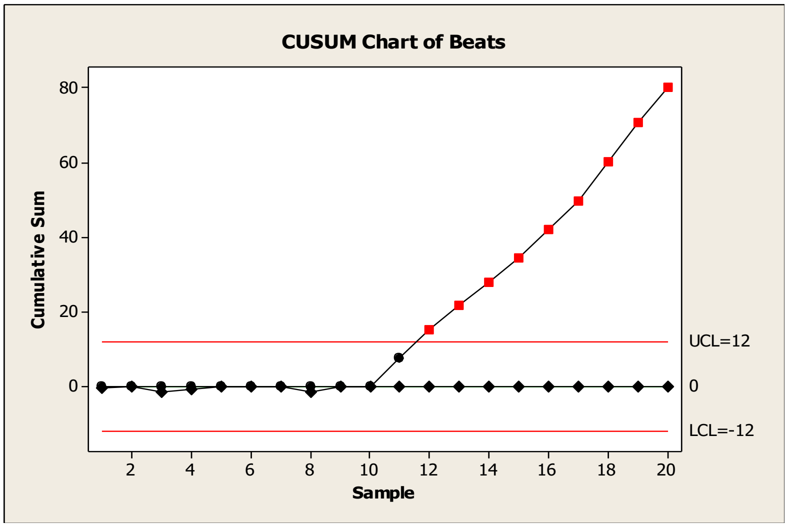
15-83 Heart rate (in counts/minute) is measured every 30 minutes. The results of 20 consecutive measurements are as follows:

Sample	Heart Rate	Sample	Heart Rate
1	68	11	79
2	71	12	79
3	67	13	78
4	69	14	78
5	71	15	78
6	70	16	79
7	69	17	79
8	67	18	82
9	70	19	82
10	70	20	81

Suppose that the standard deviation of the heart rate is $\sigma = 3$ and the target value is 70.

(a) Design a CUSUM scheme for the heart rate process. Does the process appear to be in control at the target?
(b) How many samples on average would be required to detect a shift of the mean heart rate to 80?

(a) The process is not in control. The control chart follows with h = 4 and k = 0.5.



(b) The shift from 70 to 80 is a shift of 10/3 = 3.33 standard deviation units.

Table 15-9 can be used.

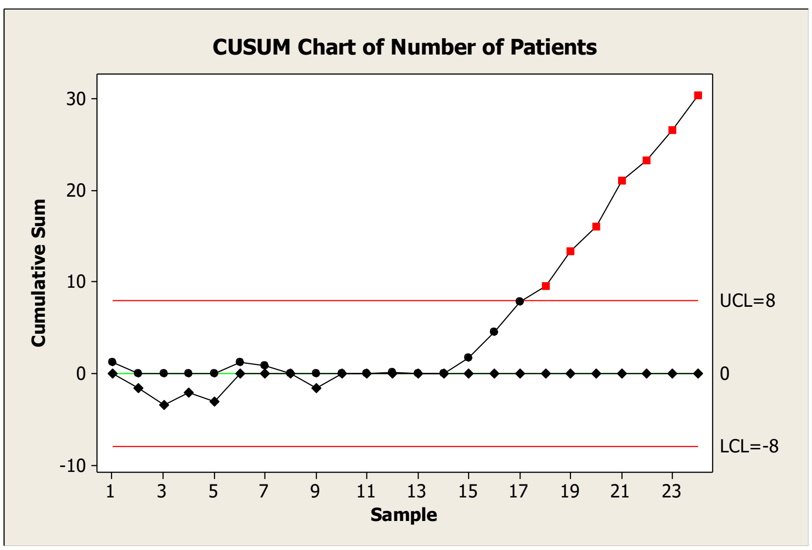
If h = 4 and k = 0.5 then the shift of 3.33 is between the table values 3 and 4. Therefore, 1.71 < ARL < 2.19.
If h = 5 and k = 0.5 then the shift of 3.33 is between the table values 3 and 4. Therefore, 2.01 < ARL < 2.57.

15-84 The number of influenza patients (in thousands) visiting hospitals weekly is shown in the following table. Suppose that the standard deviation is $\sigma = 2$ and the target value is 160.

Sample	Number of Patients	Sample	Number of Patients
1	162.27	13	159.989
2	157.47	14	159.09
3	157.065	15	162.699
4	160.45	16	163.89
5	157.993	17	164.247
6	162.27	18	162.7
7	160.652	19	164.859
8	159.09	20	163.65
9	157.442	21	165.99
10	160.78	22	163.22
11	159.138	23	164.338
12	161.08	24	164.83

(a) Design a CUSUM scheme for the process. Does the process appear to be in control at the target?
(b) How many samples on average would be required to detect a shift of the mean to 165?

(a) The process is not in control. The control chart follows.



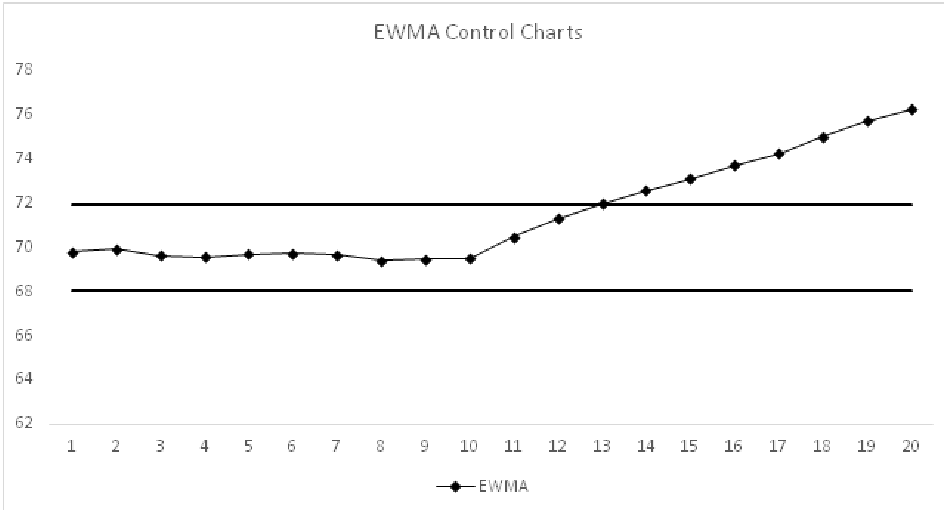
(b) The shift from 160 to 165 is a shift of 5/2 = 2.5 standard deviation units. If h = 4 and k = 0.5 then Table 15-9 can be used. The shift of 2.5 leads to ARL = 2.62

15-85 Consider the heart rate data in Exercise 15-83. Use $\mu = 70$ and $\sigma = 3$.

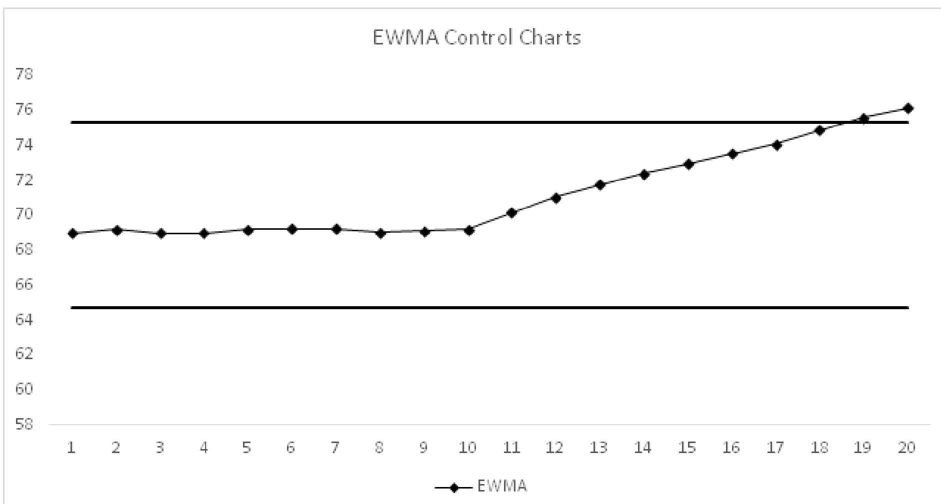
(a) Construct an EWMA control chart with $\lambda=0.1$. Use L = 2.81. Does the process appear to be in control?
(b) Construct an EWMA control chart with $\lambda=0.5$. Use L = 3.07. Compare your results to those in part (a).

(c) If the heart rate mean shifts to 76, approximate the ARLs for the charts in parts (a) and (b).

(a)$UCL = 71.93$ and$LCL = 68.07$, the chart signals at observation 13



(b) $UCL = 75.32$ and $LCL = 64.68$, the chart signals at observation 19

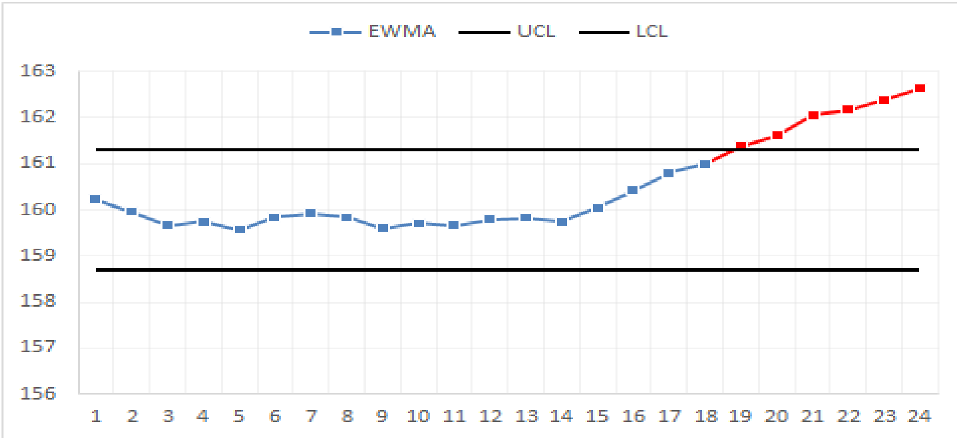


(c) A mean shift from 70 to 76, is a shift of 2σ. According to the EWMA ARL table, the ARL for part a) is 4.36 and the ARL for part b) is 3.63.

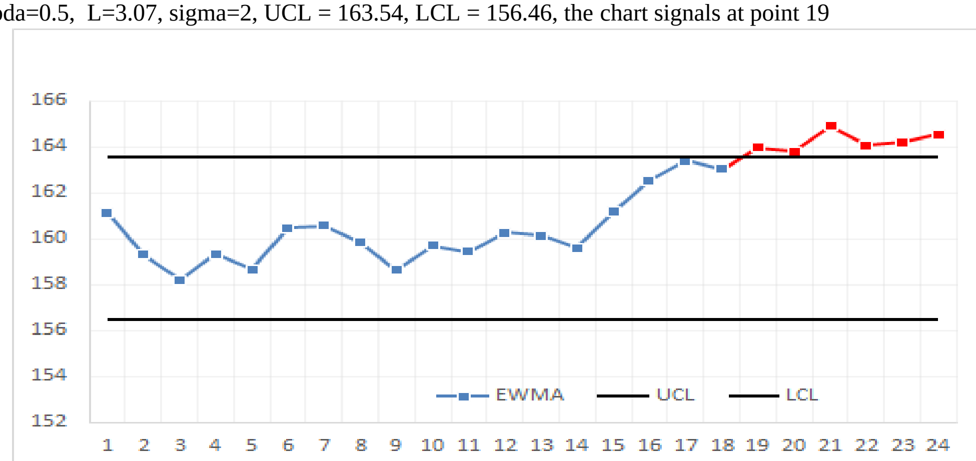
15-86 Consider the influenza data in Exercise 15-84. Use $\mu = 160$ and $\sigma = 2$.

(a) Construct an EWMA control chart with $\lambda=0.1$. Use L = 2.81. Does the process appear to be in control?
(b) Construct an EWMA control chart with $\lambda=0.5$. Use L = 3.07. Compare your results to those in part (a).

(a) lambda=0.1, L=2.81, sigma=2, UCL = 161.29, LCL = 158.71, the chart signals at point 19



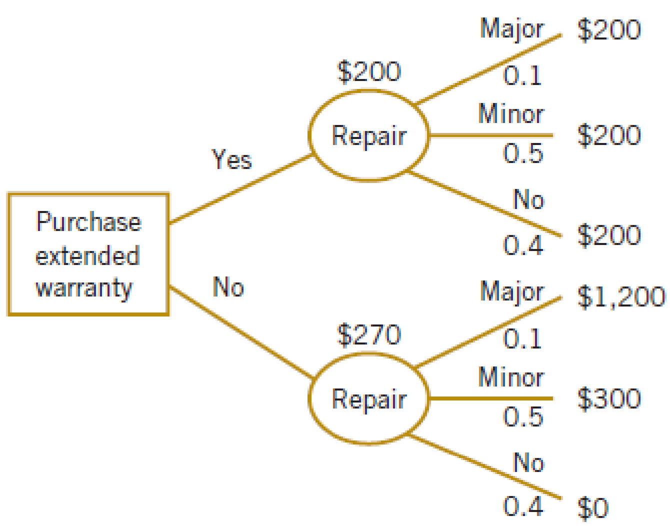
(b) lambda=0.5, L=3.07, sigma=2, UCL = 163.54, LCL = 156.46, the chart signals at point 19



The results are similar to part a).

Section 15-9

15-87 Suppose that the cost of a major repair without the extended warranty in Example 15-8 is changed to $1000. Determine the decision selected based on the minimax, most probable, and expected cost criteria.



(a) Minimax criteria: purchase cost = 200, max cost if not purchased = 1000, therefore the minimum cost decision is to purchase

(b) Most probable criteria: purchase cost = 200, most probable cost if not purchased = 300, so purchase

(c) Expected cost: purchase cost = 200, expected cost if not purchased = 0.1(1000) + 0.5(300) + 0.4(0) = 250, so purchase

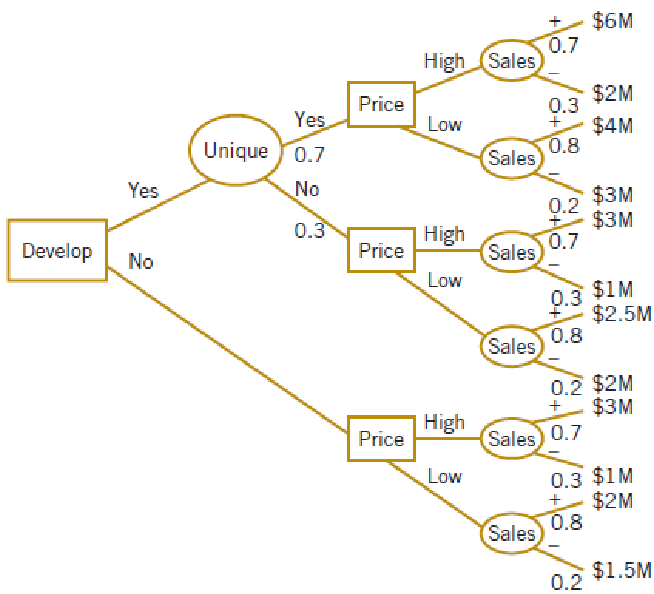
15-88 Reconsider the extended warranty decision in Example 15-8. Suppose that the probabilities of the major, minor, and no repair states are changed to 0.2, 0.4, and 0.4, respectively. Determine the decision selected based on the minimax, most probable, and expected cost criteria.

(a) Minimax criteria: purchase cost = 200, max cost if not purchased = 1200, therefore the minimum cost decision is to purchase

(b) Most probable criteria: purchase cost = 200, most probable cost if not purchased is either 300 or 0 because they both have the same probability. Because 300 is greater than 200 and 0 is less than 200, the solution from the most probable criterion is not defined in this case.

(c) Expected cost: purchase cost = 200, expected cost if not purchased = 0.2(1000) + 0.4(300) + 0.4(0) = 320, so purchase

15-89 Analyze Example 15-9 based on the most probable criterion and determine the actions that are selected at each decision node. Do any actions differ from those selected in the example?



Decisions:
1. When a new product is developed and a unique product is achieved, the most probable outcomes for the high and low prices are 6M and 4M, respectively. Therefore, the price is set high.

2. When a new product is developed and a unique product is not achieved, the most probable outcomes for the high and low prices are 3M and 2.5M, respectively. Therefore, the price is set high.

3. When a new product is not developed, the most probable outcomes for the high and low prices are 3M and 2M, respectively. Therefore, the price is set to high.

4. When a new product is developed, the most probably outcome is that it is unique. The price decision based on the most probable outcome is to price high with the most probable outcome 6M.

When a new product is not developed, the price decision based on the most probable outcome is to price high with the most probable outcome 3M.

Therefore, the decision is to develop a new product.

The choice is different from pessimistic approach in the example.

15-90 Analyze Example 15-9 based on the expected profit criterion and determine the actions that are selected at each decision node. Do any actions differ from those selected in the example?

Decisions:
1. When a new product is developed and a unique product is achieved, the expected outcomes for the high and low prices are 0.7(6M) + 0.3(2M) = 4.8M and 0.8(4M) + 0.2(3M) = 3.8, respectively. Therefore, the price is set high.

2. When a new product is developed and a unique product is not achieved, the expected outcomes for the high and low prices are 0.7(3M) + 0.3(1M) = 2.4M and 0.8(2.5M) + 0.2(2M) = 2.4, respectively. Therefore, there is no difference in expected value between the price high and low decisions.

3. When a new product is not developed, the expected outcomes for the high and low prices are 0.7(3M) + 0.3(1M) = 2.4M and 0.8(2M) + 0.2(1.5M) = 1.9M, respectively. Therefore, the price is set high.

4. When a new product is developed, the expected outcome at the unique node is 0.7(4.8M) + 0.3(2.4M) = 4.08M, where the expected outcomes at the price decision nodes are used in this calculation. When a new product is not developed, the price decision is to price high with an expected outcome of 2.4M.

Therefore, the choice from the expected value criterion is to develop a new product.

The choice is different from pessimistic approach in the example.

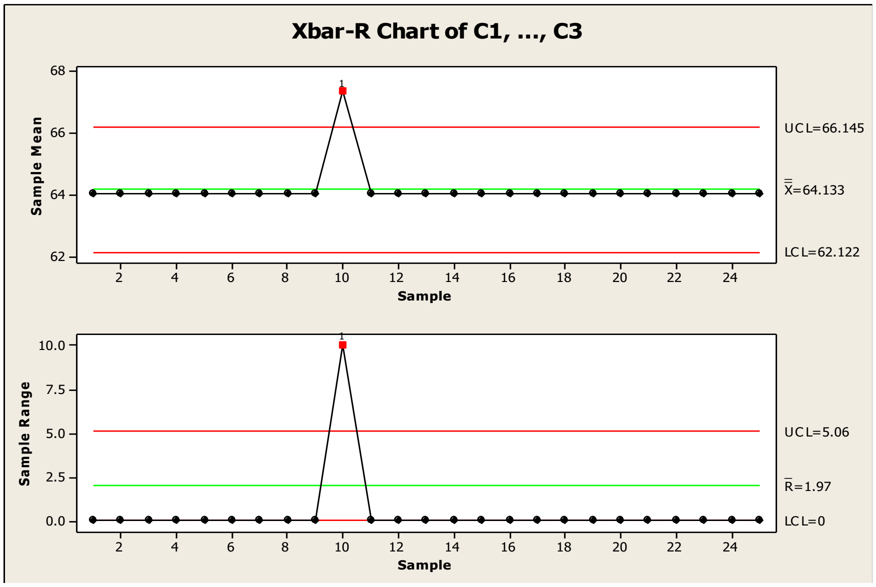
Supplementary Exercises

15-91 The diameter of fuse pins used in an aircraft engine application is an important quality characteristic. Twenty-five samples of three pins each are shown as follows:

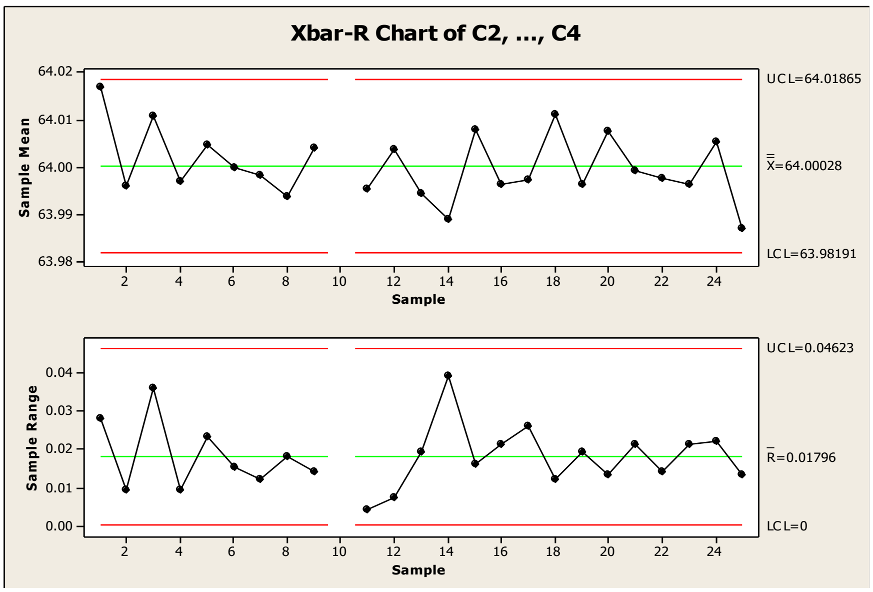
Sample Number	Diameter		
1	64.030	64.002	64.019
2	63.995	63.992	64.001
3	63.988	64.024	64.021
4	64.002	63.996	63.993
5	63.992	64.007	64.015
6	64.009	63.994	63.997
7	63.995	64.006	63.994
8	63.985	64.003	63.993
9	64.008	63.995	64.009
10	63.998	74.000	63.990
11	63.994	63.998	63.994
12	64.004	64.000	64.007
13	63.983	64.002	63.998
14	64.006	63.967	63.994
15	64.012	64.014	63.998
16	64.000	63.984	64.005
17	63.994	64.012	63.986
18	64.006	64.010	64.018
19	63.984	64.002	64.003
20	64.000	64.010	64.013
21	63.988	64.001	64.009
22	64.004	63.999	63.990
23	64.010	63.989	63.990
24	64.015	64.008	63.993
25	63.982	63.984	63.995

(a) Set up $\overline{X}$ and R charts for this process. If necessary, revise limits so that no observations are out of control.
(b) Estimate the process mean and standard deviation.
(c) Suppose that the process specifications are at 64 ± 0.02. Calculate an estimate of *PCR*. Does the process meet a minimum capability level of $PCR \geq 1.33$?
(d) Calculate an estimate of PCR_k. Use this ratio to draw conclusions about process capability.
(e) To make this process a 6-sigma process, the variance σ^2 would have to be decreased such that $PCR_k = 2.0$. What should this new variance value be?
(f) Suppose that the mean shifts to 64.01. What is the probability that this shift is detected on the next sample? What is the ARL after the shift?

(a) The process is not in control. The control chart follows.



Sample 10 is removed to obtain the following chart.



(b) estimates: mean = 64, stdev = 0.01796/1.693 = 0.0106

(c) PCR = 0.63

(d) $PCR_k = 0.63$

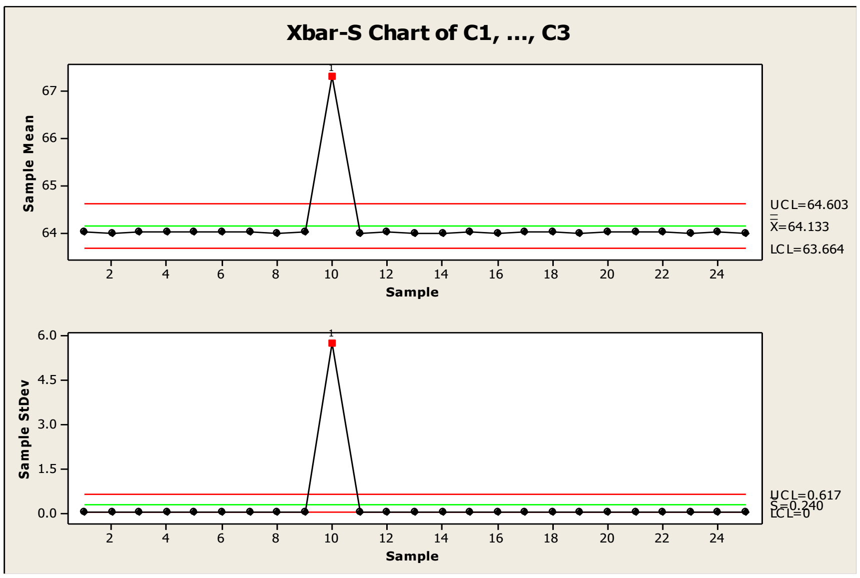
(e) The value of the variance is found by solving $PCR_k = \dfrac{\bar{\bar{x}} - LSL}{3\sigma} = 2.0$ for σ. This yields

$$\frac{64 - 63.98}{3\sigma} = 2.0 \text{, and } \sigma = 0.0033$$

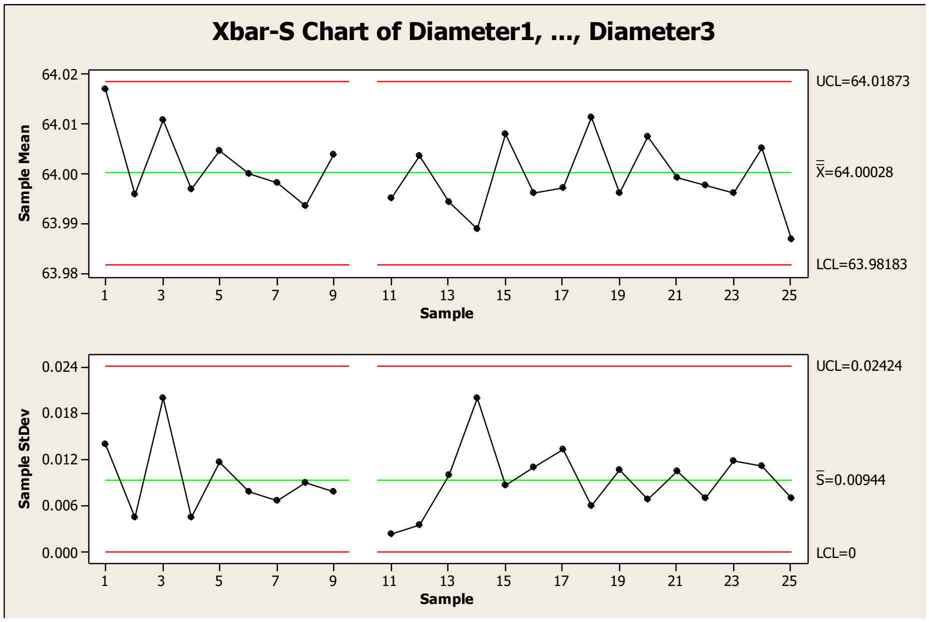
(f) P(63.9819 < Z < 64.0187)
= P[(63.9819 - 64.01)/(0.0106/sqrt(3))] < Z < (64.0187 - 64.01)/(0.0106/sqrt(3))]
=P(-4.59 < Z < 1.42) = 0.922. Therefore, the probability of detection = 1 – 0.922 = 0.078
ARL = 1/0.078 = 12.9

15-92 Rework Exercise 15-91 with $\overline{X}$ and *S* charts.

(a)



The following chart is obtained with subgroup 10 excluded:



(b) $\hat{\mu} = \bar{\bar{x}} = 64$, and $\hat{\sigma} = \dfrac{\bar{s}}{c_4} = \dfrac{0.00944}{0.8862} = 0.01065$

(c) Same as 15-73 $PCR = \dfrac{USL - LSL}{6\hat{\sigma}} = \dfrac{64.02 - 63.98}{6(0.01065)} = 0.626$

The process does not meet the minimum capability level of PCR $\geq$ 1.33.

(d) Same as the referenced exercise

$$PCR_k = \min\left[\frac{USL - \bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}} - LSL}{3\hat{\sigma}}\right] = \min\left[\frac{64.02 - 64}{3(0.01065)}, \frac{64 - 63.98}{3(0.01065)}\right]$$

$$= \min[0.626, 0.626] = 0.626$$

(e) Same as the referenced exercise

In order to make this process a "six-sigma process", the variance σ^2 would have to be decreased such that $PCR_k = 2.0$.

The value of the variance is found by solving $PCR_k = \dfrac{\bar{\bar{x}} - LSL}{3\sigma} = 2.0$ for σ:

$$\frac{64 - 63.98}{3\sigma} = 2.0, \text{ then } \sigma = 0.0033$$

Therefore, the process variance would have to be decreased to $\sigma^2 = (0.0033)^2 = 0.000011$.

(f)

$$P(63.98 < X < 64.02) = P\left(\frac{63.9818 - 64.01}{0.01065/\sqrt{3}} < \frac{X - \mu}{\sigma_x} < \frac{64.0187 - 64.01}{0.01065/\sqrt{3}}\right)$$

$$= P(-4.586 < Z < 1.415) = P(Z < 1.415) - P(Z < -4.879)$$

$$= 0.9215 - 0 = 0.9215$$

The probability that this shift will be detected on the next sample is $p = 1 - 0.9215 = 0.0785$

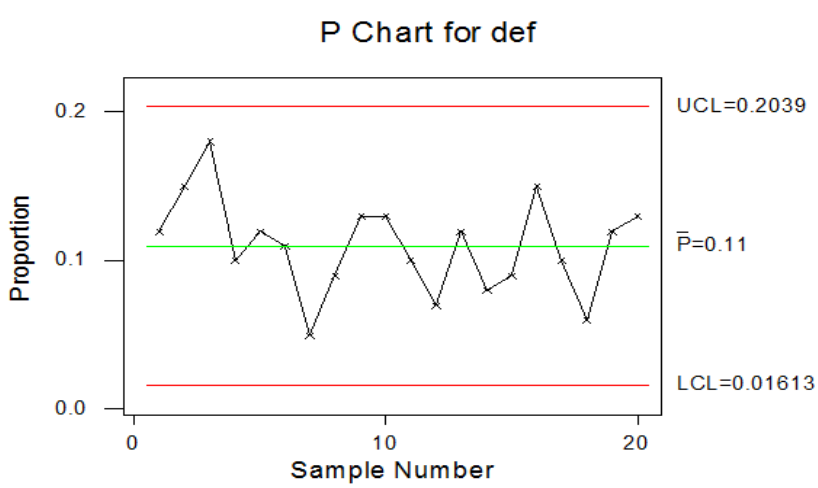
$$ARL = \frac{1}{p} = \frac{1}{0.0785} = 12.739$$

15-93 Plastic bottles for liquid laundry detergent are formed by blow molding. Twenty samples of $n = 100$ bottles are inspected in time order of production, and the fraction defective in each sample is reported. The data are as follows:

Sample	Fraction Defective
1	0.12
2	0.15
3	0.18
4	0.10
5	0.12
6	0.11
7	0.05
8	0.09
9	0.13
10	0.13
11	0.10
12	0.07
13	0.12
14	0.08
15	0.09
16	0.15
17	0.10
18	0.06
19	0.12
20	0.13

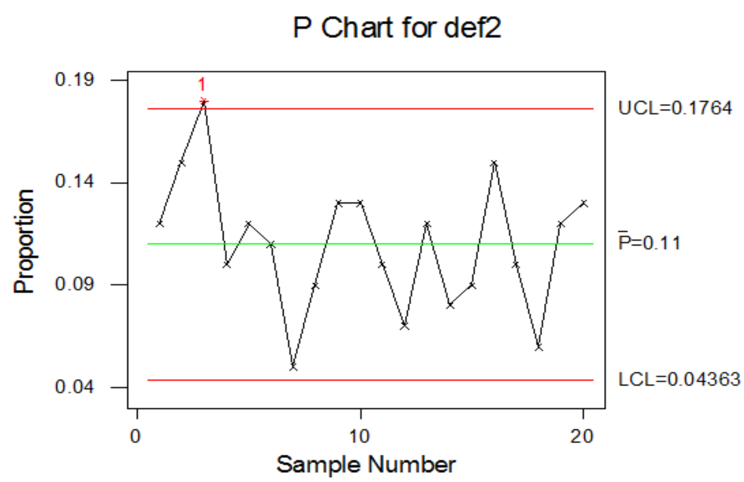
(a) Set up a *P* chart for this process. Is the process in statistical control?
(b) Suppose that instead of $n = 100$, $n = 200$. Use the data given to set up a *P* chart for this process. Revise the control limits if necessary.
(c) Compare your control limits for the *P* charts in parts (a) and
(b) Explain why they differ. Also, explain why your assessment about statistical control differs for the two sizes of *n*.

(a)

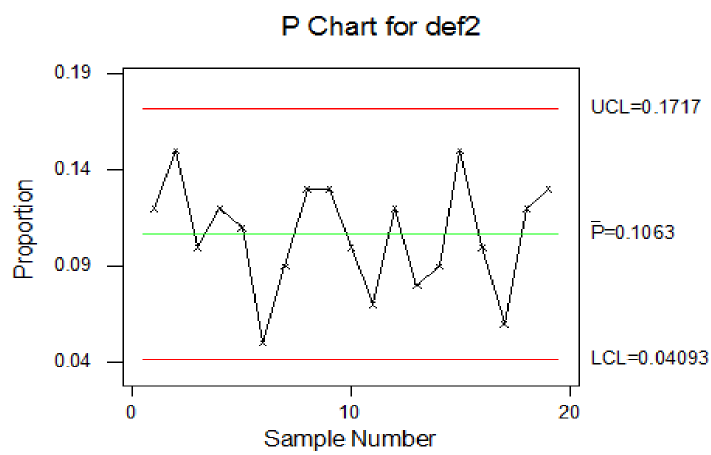


There are no points beyond the control limits. The process is in control.

(b)



There is one point beyond the upper control limit. The process is out of control. The revised limits are:



There are no further points beyond the control limits.

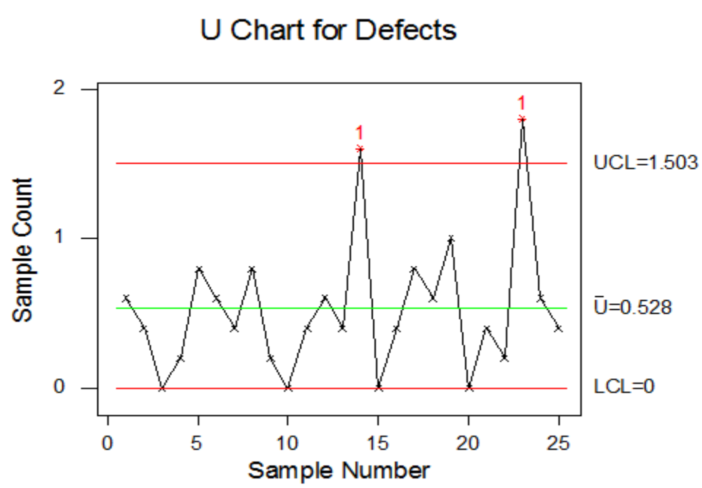
(c) A larger sample size with the same percentage of defective items will result in more narrow control limits. The control limits corresponding to the larger sample size are more sensitive to process shifts.

15-94 Cover cases for a personal computer are manufactured by injection molding. Samples of five cases are taken from the process periodically, and the number of defects is noted. Twenty-five samples follow:

Sample	No. of Defects	Sample	No. of Defects
1	3	14	8
2	2	15	0
3	0	16	2
4	1	17	4
5	4	18	3
6	3	19	5
7	2	20	0
8	4	21	2
9	1	22	1
10	0	23	9
11	2	24	3
12	3	25	2

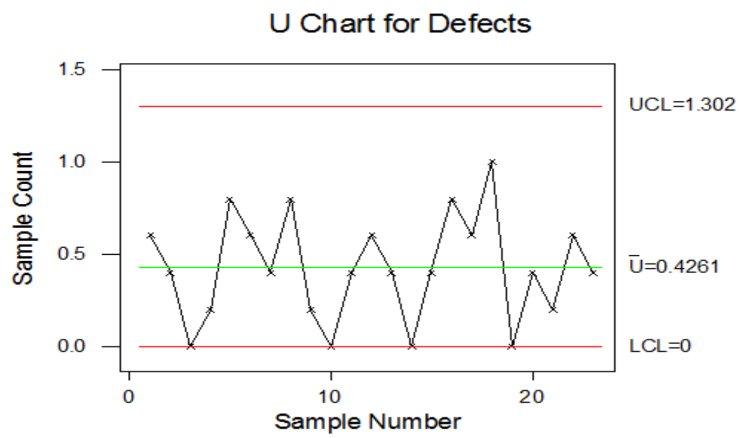
(a) Using all the data, find trial control limits for a *U* chart for the process.
(b) Use the trial control limits from part (a) to identify out-of-control points. If necessary, revise your control limits.
(c) Suppose that instead of samples of five cases, the sample size was 10. Repeat parts (a) and (b). Explain how this change alters your responses to parts (a) and (b).

(a)

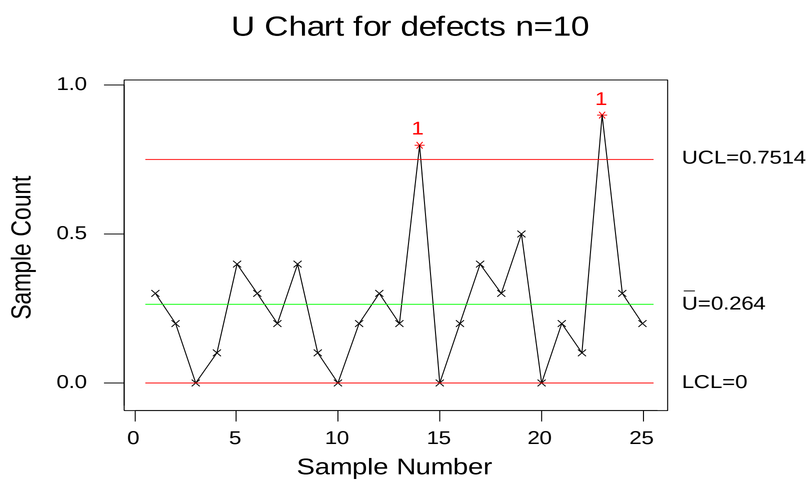


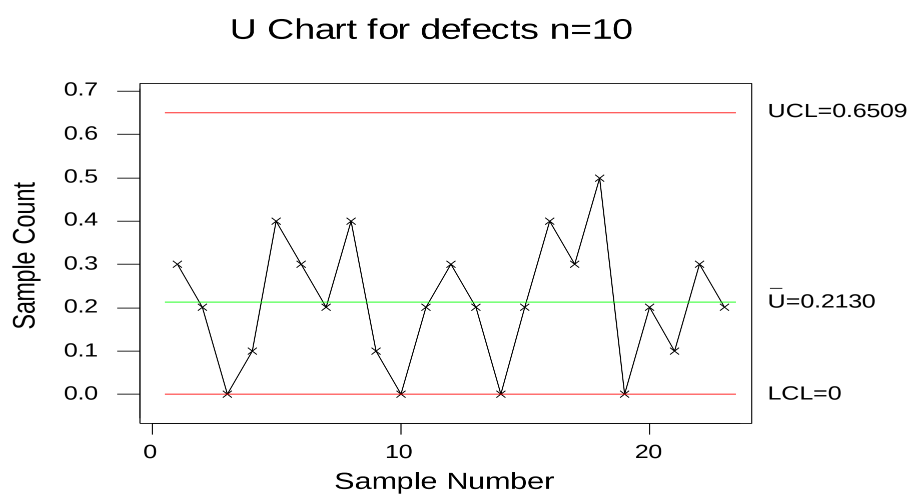
Points14 and 23 are beyond the control limits. The process is out of control.

(b) After removing points 14 and 23, the limits are narrowed.



(c) The control limits are narrowerfor a sample size of 10





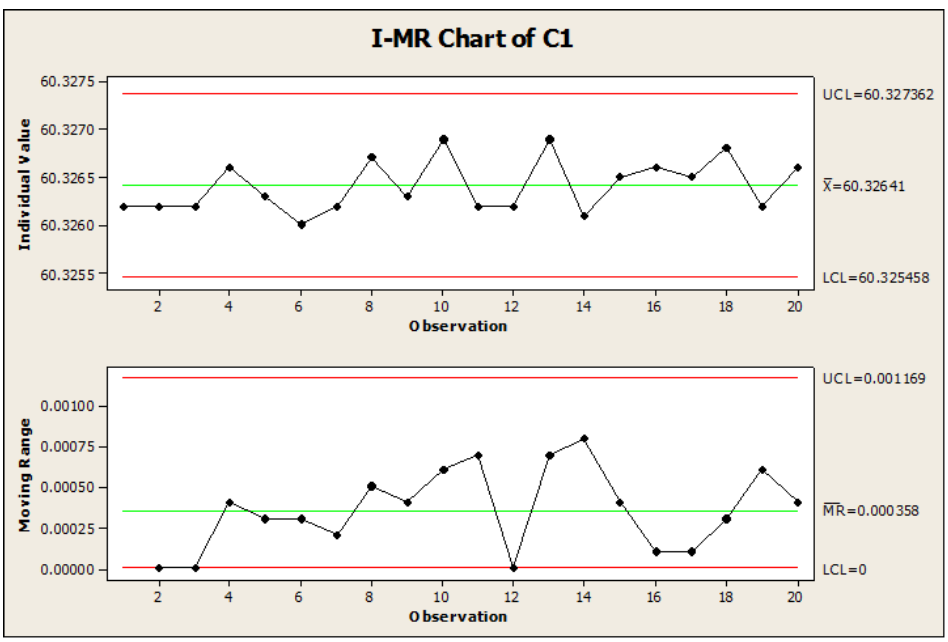
15-95 An article in *Quality Engineering* ["Is the Process Capable? Tables and Graphs in Assessing C_{pm}" (1992, Vol. 4(4)].

Considered manufacturing data. Specifications for the outer diameter of the hubs were 60.3265 ± 0.001 mm. A random sample with size n = 20 was taken and the data are shown in the following table:

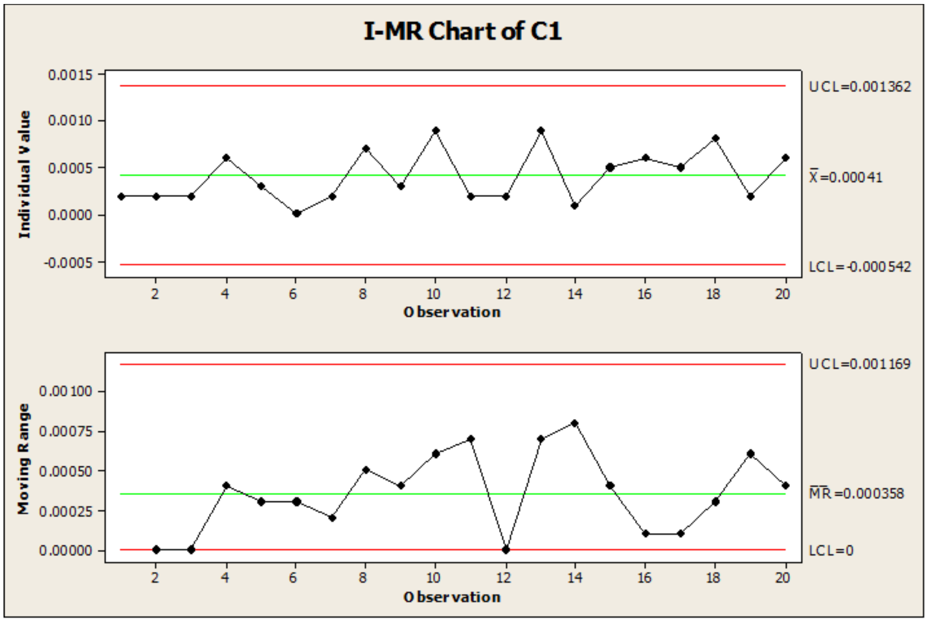
Sample	x	Sample	x
1	60.3262	11	60.3262
2	60.3262	12	60.3262
3	60.3262	13	60.3269
4	60.3266	14	60.3261
5	60.3263	15	60.3265
6	60.3260	16	60.3266
7	60.3262	17	60.3265
8	60.3267	18	60.3268
9	60.3263	19	60.3262
10	60.3269	20	60.3266

(a) Construct a control chart for individual measurements. Revise the control limits if necessary.
(b) Compare your chart in part (a) to one that uses only the last (least significant) digit of each diameter as the measurement. Explain your conclusion.
(c) Estimate μ and σ from the moving range of the revised chart and use this value to estimate *PCR* and PCR_k and interpret these ratios.

(a) Using I-MR chart.



(b) The chart is identical to the chart in part (a) except for the scale of the individuals chart.



(c)The estimated mean is 60.3264. The estimated standard deviation is 0.0003173.

$$PCR = \frac{USL - LSL}{6\sigma} = \frac{0.002}{6(0.0003173)} = 1.0505$$

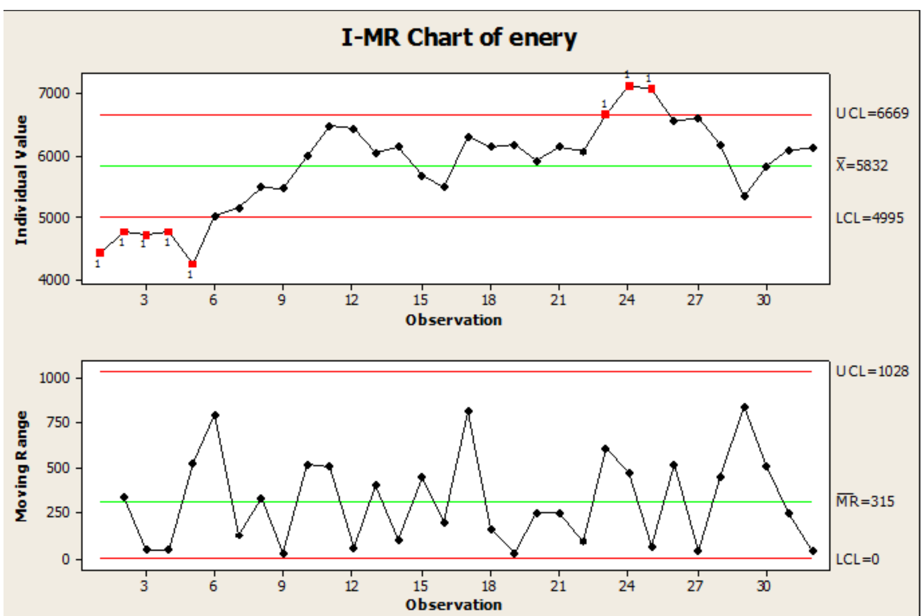
$$PCR_k = \min\left[\frac{0.0009}{3\sigma}, \frac{0.0011}{3\sigma}\right] = 0.9455$$

15-96 The following data from the U.S. Department of Energy Web site (www.eia.doe.gov) reported the total U.S. *renewable* energy consumption by year (trillion BTU) from 1973 to 2004.

Year	Total Renewable Energy Consumption (Trillion BTU)	Year	Total Renewable Energy Consumption (Trillion BTU)
1973	4433.121	1989	6294.209
1974	4769.395	1990	6132.572
1975	4723.494	1991	6158.087
1976	4767.792	1992	5907.147
1977	4249.002	1993	6155.959
1978	5038.938	1994	6064.779
1979	5166.379	1995	6669.261
1980	5494.420	1996	7136.799
1981	5470.574	1997	7075.152
1982	5985.352	1998	6560.632
1983	6487.898	1999	6598.630
1984	6430.646	2000	6158.232
1985	6032.728	2001	5328.335
1986	6131.542	2002	5835.339
1987	5686.932	2003	6081.722
1988	5488.649	2004	6116.287

(a) Using all the data, find calculate control limits for a control chart for individual measurements, construct the chart, and plot the data.
(b) Do the data appear to be generated from an in-control process? Comment on any patterns on the chart.

(a)



(b) The data does not appear to be generated from an in-control process. The average tends to drift to larger values and then drop back off over the last 5 values.

15-97 The following dataset was considered in *Quality Engineering* ["Analytic Examination of Variance Components" (1994–1995, Vol. 7(2)]. A quality characteristic for cement mortar briquettes was monitored. Samples of size $n = 6$ were taken from the process, and 25 samples from the process are shown in the following table:

(a) Using all the data, calculate trial control limits for X and S charts. Is the process in control?

Batch	$\overline{X}$	s
1	572.00	73.25
2	583.83	79.30
3	720.50	86.44
4	368.67	98.62
5	374.00	92.36
6	580.33	93.50
7	388.33	110.23
8	559.33	74.79
9	562.00	76.53
10	729.00	49.80
11	469.00	40.52
12	566.67	113.82
13	578.33	58.03
14	485.67	103.33
15	746.33	107.88
16	436.33	98.69
17	556.83	99.25
18	390.33	17.35
19	562.33	75.69
20	675.00	90.10
21	416.50	89.27
22	568.33	61.36
23	762.67	105.94
24	786.17	65.05
25	530.67	99.42

(b) Suppose that the specifications are at 580 ± 250. What statements can you make about process capability? Compute estimates of the appropriate process capability ratios.

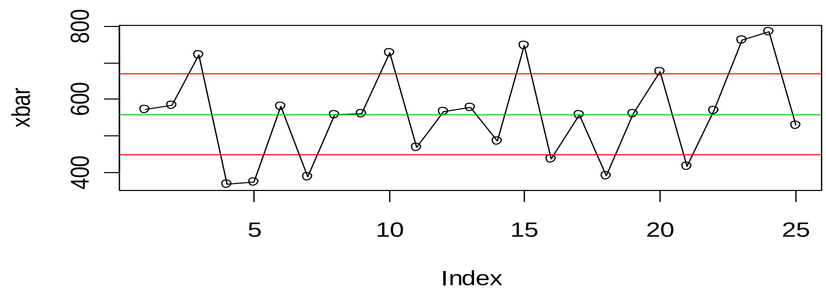
(c) To make this process a "6-sigma process," the variance σ^2 would have to be decreased such that $PCR_k = 2.0$. What should this new variance value be?

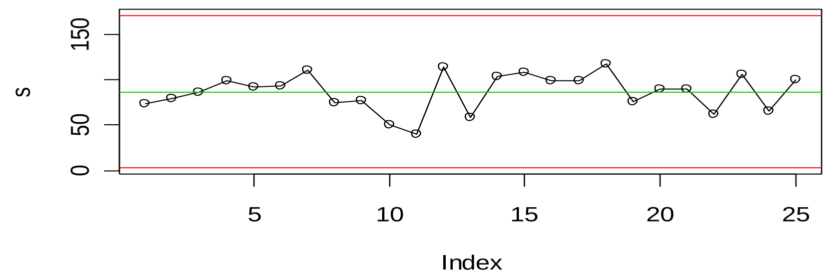
(d) Suppose the mean shifts to 600. What is the probability that this shift is detected on the next sample? What is the ARL after the shift?

(a)Trial control limits :
S chart
UCL= 170.2482
CL = 86.4208

LCL = 2.59342

X bar chart
UCL= 670.0045
CL = 558.766
LCL = 447.5275





(b) An estimate of σ is given by $\bar{S}/c_4 = 86.4208/0.9515 = 90.8259$

PCR=500/(6*90.8259)=0.9175 and PCR_k= $\min\left[\dfrac{830-558.77}{3(90.8259)}, \dfrac{558.77-330}{3(90.8259)}\right]$=0.8396

Based on the capability ratios above (both <1), the process is operating off-center and will result in a large number of non-conforming units.

(c) To determine the new variance, solve $PCR_k = 2$ for σ.

Because PCR_k= $\dfrac{558.77-330}{3\sigma}$, we find σ=38.128 or σ^2=1453.77.

(d) The probability that $\bar{X}$ falls within the control limits is

$$P\left(447.5275 < \bar{X} < 670.0045\right) = P\left(\frac{447.5275-600}{\frac{90.8259}{\sqrt{6}}} < Z < \frac{670.0045-600}{\frac{90.8259}{\sqrt{6}}}\right) =$$

$$P(-4.11 < Z < 1.89) = 0.9706$$

Thus, p=0.0294 and ARL=1/p=34.01. The probability that the shift will be detected in the nextsample is 0.0294.

For part b) through d), if we remove the out-of-control samples in part a) and recalculate the control limits, we will get the following control limits for S chart and Xbar chart.

S chart
UCL= 158.9313
CL = 80.67615
LCL = 2.421028

X bar chart
UCL= 655.7918
CL = 551.9477
LCL = 448.1035

(b) $\hat{\sigma}$ =84.78839

$$\begin{aligned} PCR_K &= \min\left[\frac{USL-\bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}}-LSL}{3\hat{\sigma}}\right] \\ &= \min\left[\frac{830-551.95}{3(84.79)}, \frac{551.95-330}{3(84.79)}\right] \\ &= 0.8725 \end{aligned}$$

(c)

$$\begin{aligned} PCR_K &= \min\left[\frac{USL-\bar{\bar{x}}}{3\hat{\sigma}}, \frac{\bar{\bar{x}}-LSL}{3\hat{\sigma}}\right] \\ &= \frac{551.95-330}{3(\hat{\sigma})} \\ &= 2 \end{aligned}$$

So $\hat{\sigma} = 36.9917$

(d)In-control distribution $\overline{X}$ ~ N(551.9477, 34.6^2)
Out-of-control distribution $\overline{X}$ ~ N(600, 34.6^2)

P[448.1 $\le \overline{X} \le$ 655.8, when μ =600]

$= P[\frac{448.1-600}{34.6} \le Z \le \frac{655.8-600}{34.6}]$

$= P[-4.39 \le Z \le 1.6124]$

=0.9463

Out-of-control ARL= $\frac{1}{1-0.9463} = 18.6 \approx 19.$

15-98 Suppose that an $\overline{X}$ control chart with 2-sigma limits is used to control a process. Find the probability that a false out-of-control signal is produced on the next sample. Compare this with the corresponding probability for the chart with 3-sigma limits and discuss. Comment on when you would prefer to use 2-sigma limits instead of 3-sigma limits.

$\overline{\text{X}}$ Control Chart with 2-sigma limits

$CL = \mu$

$UCL = \mu + 2\frac{\sigma}{\sqrt{n}}, LCL = \mu - 2\frac{\sigma}{\sqrt{n}}$

$$P\left(\overline{X} > \mu + 2\frac{\sigma}{\sqrt{n}}\right) = P\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} > 2\right) = P(Z > 2) = 1 - P(Z < 2) = 1 - 0.97725 = 0.02275$$

and

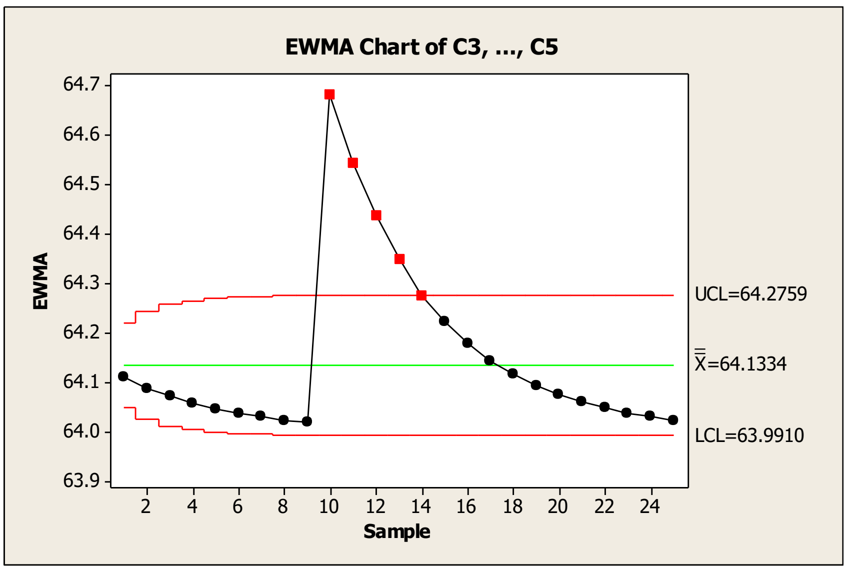
$$P\left(\overline{X} < \mu - 2\frac{\sigma}{\sqrt{n}}\right) = P\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < -2\right) = P(Z < -2) = 1 - P(Z < 2) = 1 - 0.97725 = 0.02275$$

The answer is 0.02275 + 0.02275 = 0.0455. The answer for 3-sigma control limits is 0.0027. The 3-sigma control limits result in many fewer false alarms.

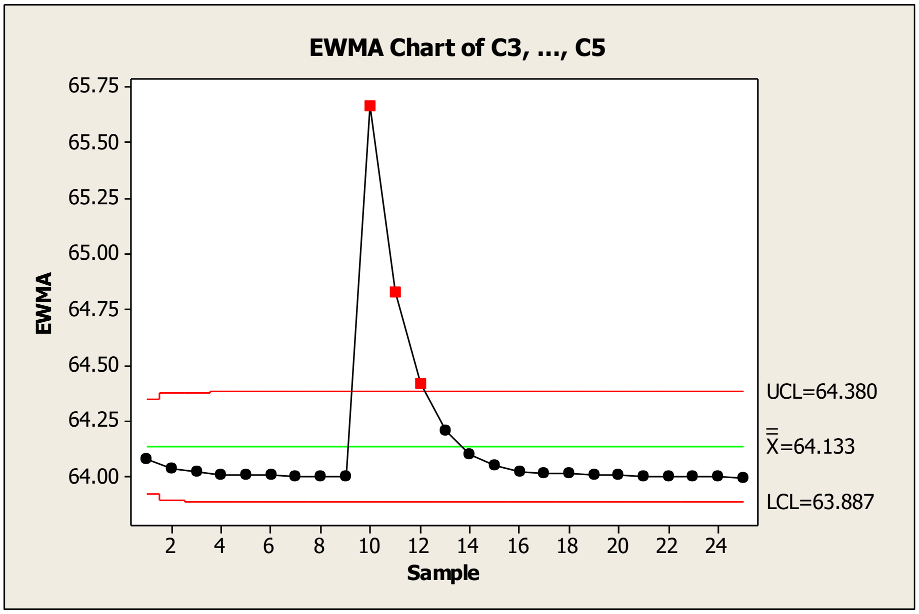
15-99 Consider the diameter data in Exercise 15-91.

(a) Construct an EWMA control chart with $\lambda = 0.2$ and $L = 3$. Comment on process control.
(b) Construct an EWMA control chart with $\lambda = 0.5$ and $L = 3$ and compare your conclusion to part (a).

(a) The following control chart use the average range from 25 subgroups of size 3 to estimate the process standard deviation. The software uses a pooled estimate of variance as the default method for an EWMA control chart so that the range method was selected from the options. Points are clearly out of control.



(b) The following control chart use the average range from 25 subgroups of size 3 to estimate the process standard deviation. There is a large shift in the mean at sample 10 and the process is out of control at this point.

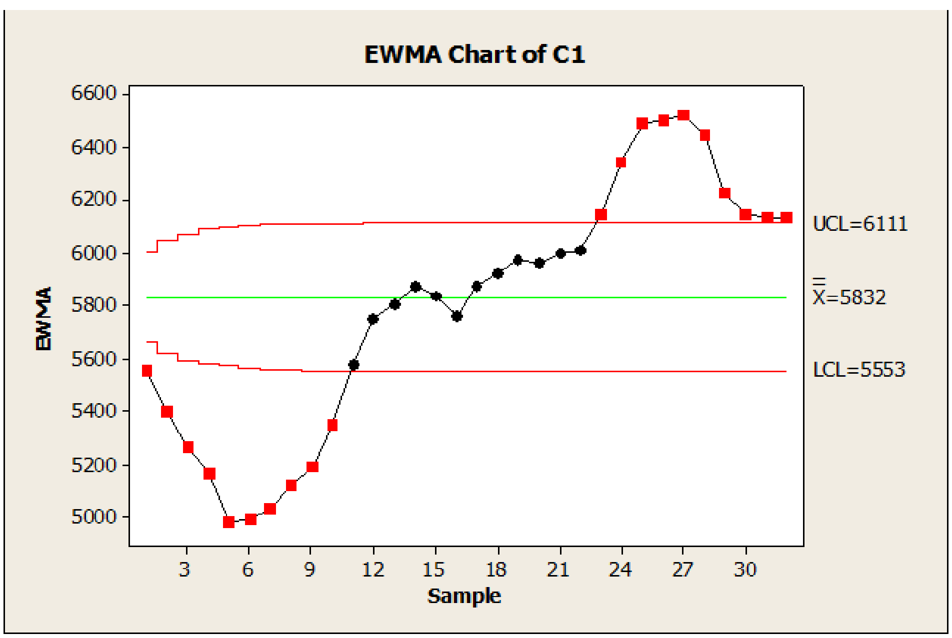


15-100 Consider the renewable energy data in Exercise 15-96.

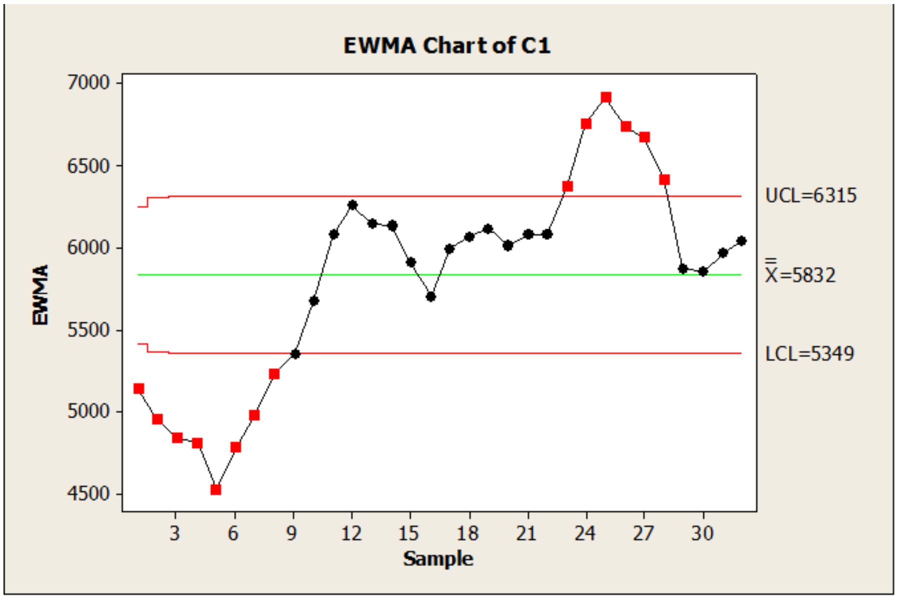
(a) Construct an EWMA control chart with $\lambda = 0.2$ and L = 3. Do the data appear to be generated from an incontrol process?

(b) Construct an EWMA control chart with $\lambda = 0.5$ and L = 3 and compare your conclusion to part (a).

(a) The data appear to be generated from an out-of-control process.



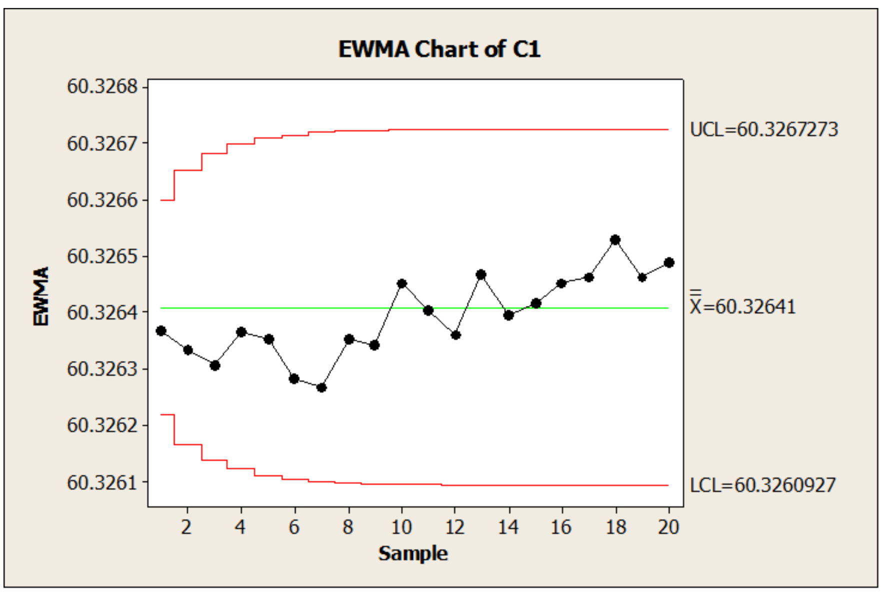
(b) The data appear to be generated from an out-of-control process.



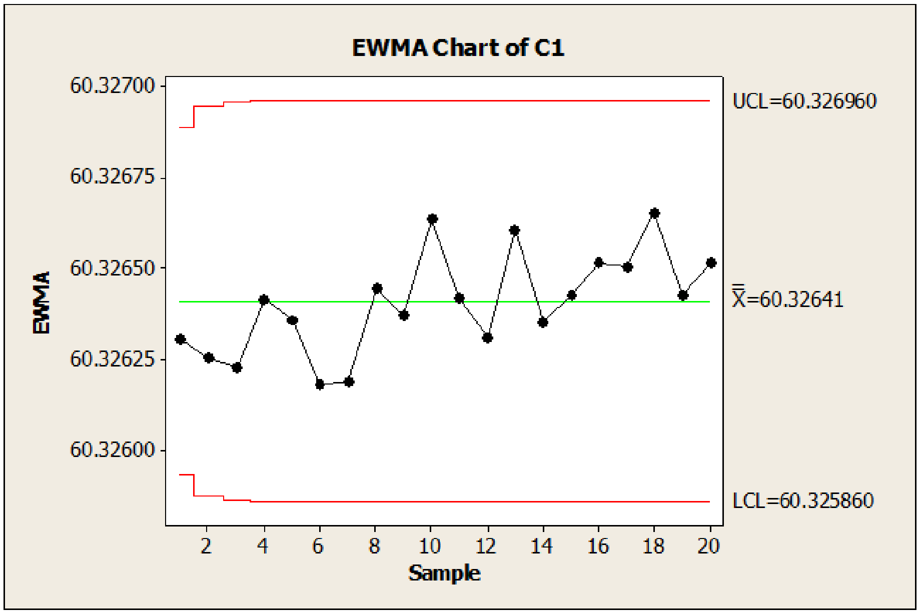
15-101 Consider the hub data in Exercise 15-95.

(a) Construct an EWMA control chart with $\lambda = 0.2$ and $L = 3$. Comment on process control.
(b) Construct an EWMA control chart with $\lambda = 0.5$ and $L = 3$ and compare your conclusion to part (a).

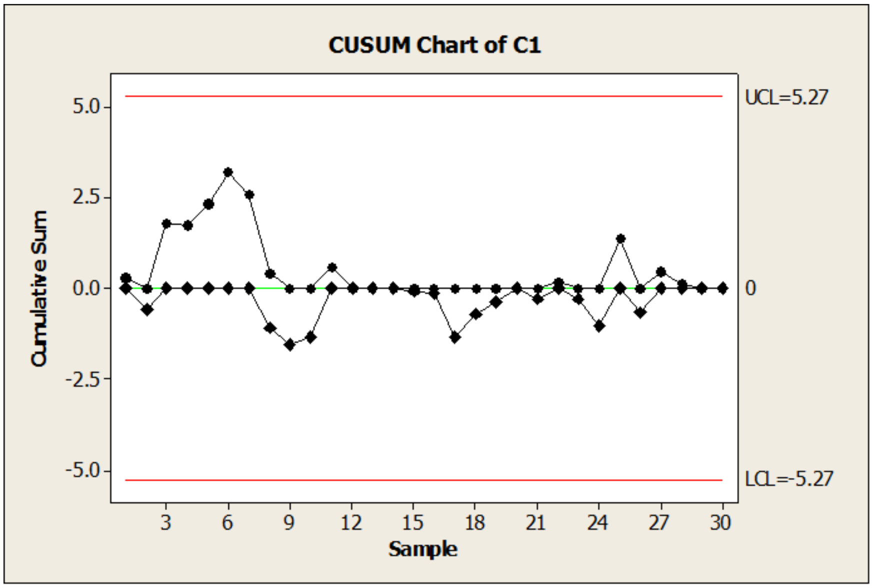
(a) The process appears to be in control.



(b) The process appears to be in control.



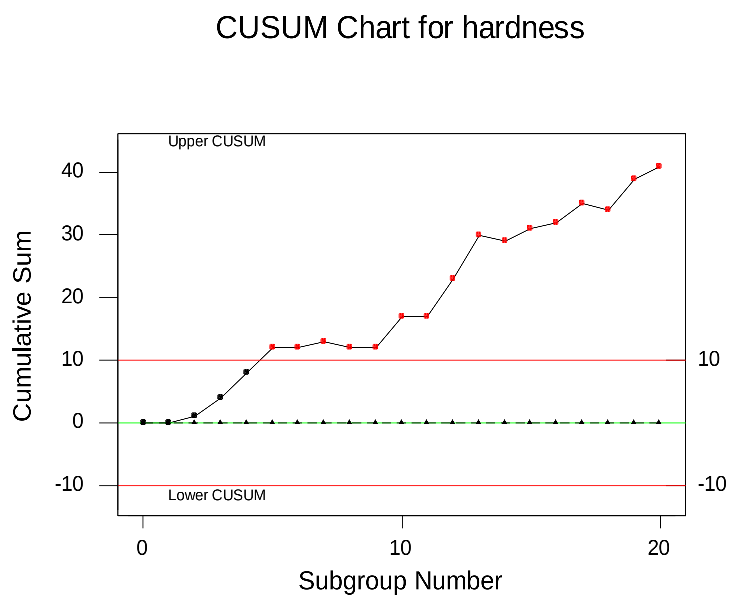
15-102 Consider the data in Exercise 15-20. Set up a CUSUM scheme for this process assuming that $\mu = 16$ is the process target. Explain how you determined your estimate of σ and the CUSUM parameters K and H.



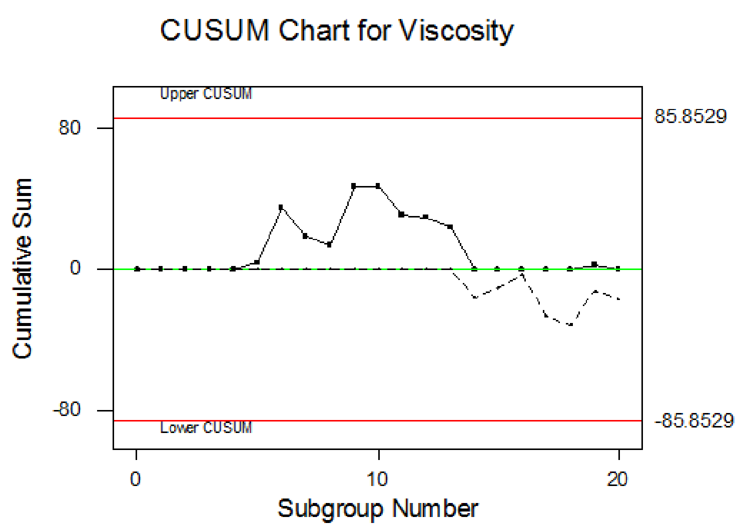
The process standard deviation is estimated using the average moving range of size 2 with MR/d_2, where $d_2 = 1.128$. The estimate is 1.05. Recommendation for k and h are 0.5 and 4 or 5, respectively for n =1. For this chart $h = 5$ was used.

15-103 Consider the hardness measurement data in Exercise 15-19. Set up a CUSUM scheme for this process using $\mu = 50$ and $\sigma = 2$ so that $K = 1$ and $H = 10$. Is the process in control?

The process is not in control.
$K = k\sigma = 1$, so that $k = 0.5$
$H = h\sigma = 10$, so that $h = 5$



15-104 Reconsider the viscosity data in Exercise 15-22. Construct a CUSUM control chart for this process using $\mu_0 = 500$ as the process target. Explain how you determined your estimate of σ and the CUSUM parameters *H* and *K*.



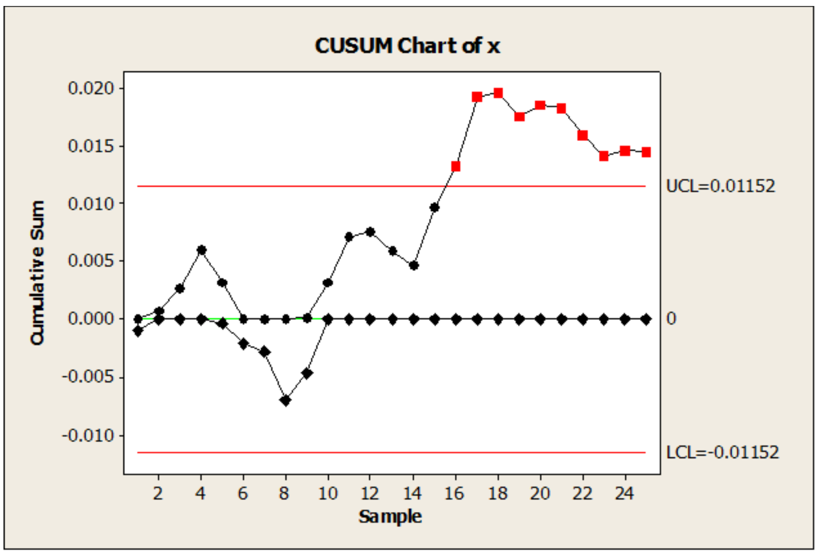
Process standard deviation is estimated using the average moving range of size 2 with MR/d_2, where $d_2 = 1.128$ for a moving range of 2. The estimate is 17.17. Recommendation for *k* and *h* are 0.5 and 4 or 5, respectively, for $n = 1$.

15-105 The following data were considered in *Quality Progress* ["Digidot Plots for Process Surveillance" (1990, May, pp. 66–68)]. Measurements of center thickness (in mils) from 25 contact lenses sampled from the production process at regular intervals are shown in the following table.

Sample	x	Sample	x
1	0.3978	14	0.3999
2	0.4019	15	0.4062
3	0.4031	16	0.4048
4	0.4044	17	0.4071
5	0.3984	18	0.4015
6	0.3972	19	0.3991
7	0.3981	20	0.4021
8	0.3947	21	0.4009
9	0.4012	22	0.3988
10	0.4043	23	0.3994
11	0.4051	24	0.4016
12	0.4016	25	0.4010
13	0.3994		

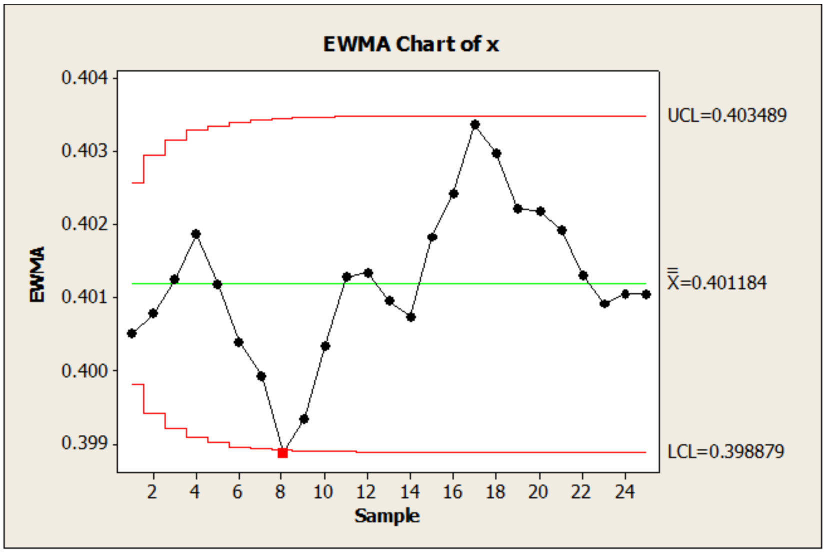
(a) Construct a CUSUM scheme for this process with the target μ_0 = 0.4. Explain how you determined your estimate of σ and the CUSUM parameters *H* and *K*. Is the process in control?
(b) Construct an EWMA control chart with $\lambda = 0.5$ and *L* = 3 and compare your conclusions to part (a).

(a)



Here σ is estimated using the moving range: 0.0026/1.128=0.0023. H and K were computed using k = 0.5 and h = 5. The process is not in control.

(b) EWMA gives similar results.



15-106 Suppose that a process is in control and an $\overline{X}$ chart is used with a sample size of 4 to monitor the process. Suddenly there is a mean shift of 1.5σ.

(a) If 3-sigma control limits are used on the $\overline{X}$ chart, what is the probability that this shift remains undetected for three consecutive samples?
(b) If 2-sigma control limits are in use on the $\overline{X}$ chart, what is the probability that this shift remains undetected for three consecutive samples?
(c) Compare your answers to parts (a) and (b) and explain why they differ. Also, which limits you would recommend using and why?

(a) Let *p* denote the probability that a point plots outside of the control limits when the mean has shifted from μ_0 to $\mu = \mu_0 + 1.5\sigma$. Then,

$$
\begin{aligned}
P(LCL < \overline{X} < UCL) &= P\left(\mu_0 - \frac{3\sigma}{\sqrt{n}} < \overline{X} < \mu_0 + \frac{3\sigma}{\sqrt{n}}\right) \\
&= P\left(\frac{-1.5\sigma}{\sigma/\sqrt{n}} - 3 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{1.5\sigma}{\sigma/\sqrt{n}} + 3\right) \\
&= P(-6 < Z < 0) = P(Z < 0) - P(Z < -6) \\
&= 0.5 - [0] = 0.5
\end{aligned}
$$

Therefore, the probability the shift is undetected for three consecutive samples is $(1-p)^3 = (0.5)^3 = 0.125$.

(b) If 2-sigma control limits were used, then

$$
\begin{aligned}
1 - p = P(LCL < \overline{X} < UCL) &= P\left(\mu_0 - \frac{2\sigma}{\sqrt{n}} < \overline{X} < \mu_0 + \frac{2\sigma}{\sqrt{n}}\right) \\
&= P\left(\frac{-1.5\sigma}{\sigma/\sqrt{n}} - 2 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{1.5\sigma}{\sigma/\sqrt{n}} + 2\right) \\
&= P(-5 < Z < -1) = P(Z < -1) - P(Z < -5) \\
&= 1 - 0.84134 - [0] = 0.15866
\end{aligned}
$$

Therefore, the probability the shift is undetected for three consecutive samples is $(1-p)^3 = (0.15866)^3 = 0.004$.

(c) The 2-sigma limits are narrower than the 3-sigma limits. Because the 2-sigma limits have a smallerprobability of a shift being undetected, the 2-sigma limits would be better than the 3-sigma limits for a mean shift of 1.5σ. However, the 2-sigma limits would result in more signals when the process has not shifted (false alarms).

15-107 Consider the control chart for individuals with 3-sigma limits.

(a) Suppose that a shift in the process mean of magnitude σ occurs. Verify that the ARL for detecting the shift is ARL = 43.9.
(b) Find the ARL for detecting a shift of magnitude 2σ in the process mean.
(c) Find the ARL for detecting a shift of magnitude 3σ in the process mean.
(d) Compare your responses to parts (a), (b), and (c) and explain why the ARL for detection is decreasing as the magnitude of the shift increases.

ARL = 1/*p* where *p* is the probability a point falls outside the control limits.
(a) $\mu = \mu_0 + \sigma$ and n = 1

$$\begin{aligned}
p &= P(\overline{X} > UCL) + P(\overline{X} < LCL) \\
&= P\left(Z > \frac{\mu_0 + \frac{3\sigma}{\sqrt{n}} - \mu_0 - \sigma}{\sigma/\sqrt{n}} \right) + P\left(Z < \frac{\mu_0 - \frac{3\sigma}{\sqrt{n}} - \mu_0 - \sigma}{\sigma/\sqrt{n}} \right) \\
&= P(Z > 3 - \sqrt{n}) + P(Z < -3 - \sqrt{n}) \\
&= P(Z > 2) + P(Z < -4) \qquad \textit{when} \quad n = 1 \\
&= 1 - 0.97725 + [0] \\
&= 0.02275
\end{aligned}$$

Therefore, ARL = 1/*p* = 1/0.02275 = 43.9.

(b) $\mu = \mu_0 + 2\sigma$

$$\begin{aligned}
&P(\overline{X} > UCL) + P(\overline{X} < LCL) \\
&= P\left(Z > \frac{\mu_0 + \frac{3\sigma}{\sqrt{n}} - \mu_0 - 2\sigma}{\sigma/\sqrt{n}} \right) + P\left(Z < \frac{\mu_0 - \frac{3\sigma}{\sqrt{n}} - \mu_0 - 2\sigma}{\sigma/\sqrt{n}} \right) \\
&= P(Z > 3 - 2\sqrt{n}) + P(Z < -3 - 2\sqrt{n}) \\
&= P(Z > 1) + P(Z < -5) \qquad \textit{when} \quad n = 1 \\
&= 1 - 0.84134 + [0] \\
&= 0.15866
\end{aligned}$$

Therefore, ARL = 1/*p* = 1/0.15866 = 6.30.

(c) $\mu = \mu_0 + 3\sigma$

$$P(\overline{X} > UCL) + P(\overline{X} < LCL)$$
$$= P\left(Z > \frac{\mu_0 + \frac{3\sigma}{\sqrt{n}} - \mu_0 - 3\sigma}{\sigma / \sqrt{n}} \right) + P\left(Z < \frac{\mu_0 - \frac{3\sigma}{\sqrt{n}} - \mu_0 - 3\sigma}{\sigma / \sqrt{n}} \right)$$
$$= P(Z > 3 - 3\sqrt{n}) + P(Z < -3 - 3\sqrt{n})$$
$$= P(Z > 0) + P(Z < -6) \qquad when \quad n = 1$$
$$= 1 - 0.50 + [0]$$
$$= 0.50$$

Therefore, ARL = 1/*p* = 1/0.50 = 2.00.

(d) The ARL is decreasing as the magnitude of the shift increases from σ to 2σ to 3σ. The ARL decrease as the magnitude of the shift increases since a larger shift is more likely to be detected earlier than a smaller shift.

15-108 Consider a control chart for individuals applied to a continuous 24-hour chemical process with observations taken every hour.

(a) If the chart has 3-sigma limits, verify that the in-control ARL is 370. How many false alarms would occur each 30-day month, on the average, with this chart?
(b) Suppose that the chart has 2-sigma limits. Does this reduce the ARL for detecting a shift in the mean of magnitude σ? (Recall that the ARL for detecting this shift with 3-sigma limits is 43.9.)
(c) Find the in-control ARL if 2-sigma limits are used on the chart. How many false alarms would occur each month with this chart? Is this in-control ARL performance satisfactory? Explain your answer.

(a) Because ARL = 370, on the average we expect there to be one false alarm every 370 hours. Each 30-day

month contains 30 × 24 = 720 hours of operation. Consequently, we expect 720/370 = 1.9 false alarmseach month

$$P(X > \overline{X} + 3\hat{\sigma}) + P(X < \overline{X} - 3\hat{\sigma}) = P(z > 3) + P(z < -3) = 2(0.00135) = 0.0027$$

ARL=1/p=1/0.0027=370.37

(b) With 2-sigma limits the probability of a point plotting out of control is determined as follows, when $\mu = \mu_0 + \sigma$

$$P(X > UCL) + P(X < LCL)$$
$$= P\left(\frac{X - \mu_0 - \sigma}{\sigma} > \frac{\mu_0 + 2\sigma - \mu_0 - \sigma}{\sigma} \right) + P\left(\frac{X - \mu_0 - \sigma}{\sigma} < \frac{\mu_0 - 2\sigma - \mu_0 - \sigma}{\sigma} \right)$$
$$= P(Z > 1) + P(Z < -3)$$
$$= 1 - P(Z < 1) + [1 - P(Z < 3)]$$
$$= 1 - 0.84134 + 1 - 0.99865$$
$$= 0.160$$

Therefore, ARL=1/*p* = 1/0.160 = 6.25. The 2-sigma limits reduce the ARL for detecting a shift in the mean of magnitude σ. However, the next part of this solution shows that the number of false alarms increases with 2-sigma limits.

(c)2σ limits

$$P(X > \overline{X} + 2\hat{\sigma}) + P(X < \overline{X} - 2\hat{\sigma}) = P(z > 2) + P(z < -2) = 2(0.02275) = 0.0455$$

ARL=1/*p*=1/0.0455=21.98. This ARL is not satisfactory. There would be too many false alarms. We would expect 32.76 false alarms per month.

15-109 The depth of a keyway is an important part quality characteristic. Samples of size *n* = 5 are taken every four hours from the process, and 20 samples are summarized in the following table.

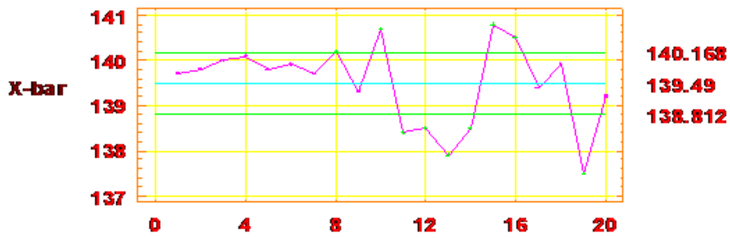
(a) Using all the data, find trial control limits for $\overline{X}$ and R charts. Is the process in control?

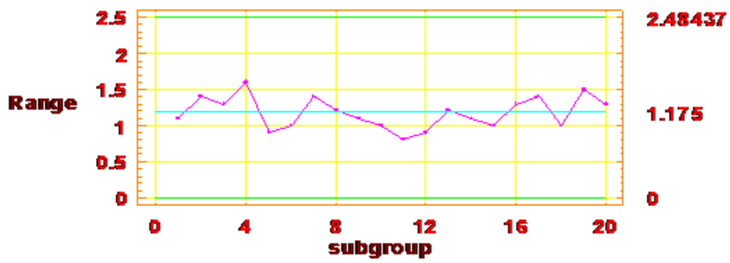
(b) Use the trial control limits from part (a) to identify outof- control points. If necessary, revise your control limits. Then estimate the process standard deviation.
(c) Suppose that the specifications are at 140 ± 2. Using the results from part (b), what statements can you make about process capability? Compute estimates of the appropriate process capability ratios.
(d) To make this a 6-sigma process, the variance s2 would have to be decreased such that PCR_k = 2.0. What should this new variance value be?
(e) Suppose that the mean shifts to 139.7. What is the probability that this shift is detected on the next sample? What is the ARL after the shift?

Sample	$\overline{X}$	r
1	139.7	1.1
2	139.8	1.4
3	140.0	1.3
4	140.1	1.6
5	139.8	0.9
6	139.9	1.0
7	139.7	1.4
8	140.2	1.2
9	139.3	1.1
10	140.7	1.0
11	138.4	0.8
12	138.5	0.9
13	137.9	1.2
14	138.5	1.1
15	140.8	1.0
16	140.5	1.3
17	139.4	1.4
18	139.9	1.0
19	137.5	1.5
20	139.2	1.3

(a)

```
        X-bar and Range - Initial Study
            Charting xbar
X-bar                          | Range
-----                          | -----
UCL:  +   3.0 sigma = 140.168  | UCL:  +   3.0 sigma = 2.48437
Centerline          = 139.49   | Centerline          = 1.175
LCL:  -   3.0 sigma = 138.812  | LCL:  -   3.0 sigma = 0
out of limits = 9              | out of limits = 0
Estimated
process mean  = 139.49
process sigma =  0.505159
mean Range    =  1.175
```

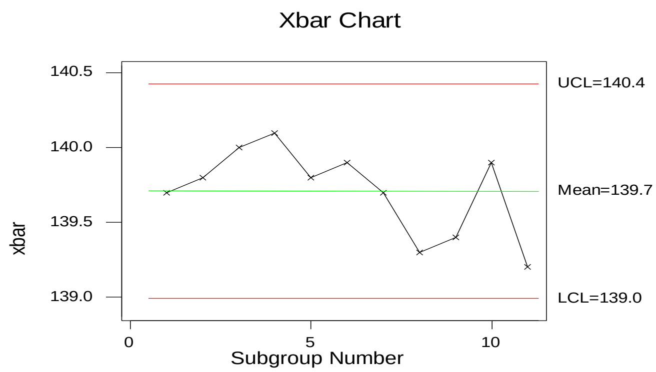


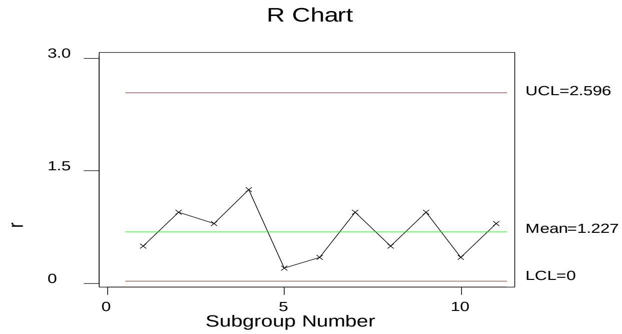


There are points beyond the control limits. The process is out of control. The points are 8, 10, 11, 12, 13,14, 15, 16, and 19.

(b) Revised control limits are given in the table below:

```
                    X-bar and Range - Initial Study
                         Charting Xbar
X-bar                          | Range
-----                          | -----
UCL:  +    3.0 sigma = 140.417| UCL:  +    3.0 sigma = 2.595682
Centerline          = 139.709| Centerline          = 1.227273
LCL:  -    3.0 sigma = 139.001        | LCL:  -    3.0 sigma = 0
out of limits = 0                | out of limits = 0
```





There are no further points beyond the control limits.

The process standard deviation estimate is given by $\hat{\sigma} = \dfrac{\overline{R}}{d_2} = \dfrac{1.227273}{2.326} = 0.5276$

(c) $PCR = \dfrac{USL - LSL}{6\hat{\sigma}} = \dfrac{142 - 138}{6(0.528)} = 1.26$

$$PCR_k = \min\left[\frac{USL - \overline{\overline{x}}}{3\hat{\sigma}}, \frac{\overline{\overline{x}} - LSL}{3\hat{\sigma}}\right] = \min\left[\frac{142 - 139.709}{3(0.528)}, \frac{139.709 - 138}{3(0.528)}\right]$$
$$= \min\left[1.45, 1.08\right] = 1.08$$

Because the process capability ratios are less than unity, the process capability appears to be poor. PCR is slightly larger than PCR_k indicating that the process is somewhat off center.

(d) In order to make this process a "six-sigma process", the variance σ^2 would have to be decreased such that PCR_k = 2.0. The value of the variance is found by solving $PCR_k = \dfrac{\overline{\overline{x}} - LSL}{3\sigma} = 2.0$ for σ:

$$\frac{139.709 - 138}{3\sigma} = 2.0$$
$$6\sigma = 139.709 - 138$$
$$\sigma = \frac{139.709 - 138}{6}$$
$$\sigma = 0.2848$$

Therefore, the process variance would have to be decreased to $\sigma^2 = (0.2848)^2 = 0.081$.

(e) $\hat{\sigma}_{\overline{x}} = 0.528$

$$p = P(139.001 < X < 140.417 \mid \mu = 139.7)$$
$$= P\left(\frac{139.001 - 139.7}{0.528} < \frac{X - \mu}{\sigma_x} < \frac{140.417 - 139.7}{0.528}\right)$$
$$= P(-1.32 < Z < 1.35)$$
$$= P(Z < 1.36) - P(Z < -1.32)$$
$$= 0.913085 - 0.093418$$
$$= 0.8197$$

The probability that this shift will be detected on the next sample is $1 - p = 1 - 0.8197 = 0.1803$.

$$ARL = \frac{1}{1-p} = \frac{1}{0.1803} = 5.55$$

15-110 Consider a control chart for individuals with 3-sigma limits. What is the probability that there is not a signal in 3 samples? In 6 samples? In 10 samples?

(a) The probability of having no signal is $P(-3 < X < 3) = 0.9973$
P(No signal in 3 samples)=$(0.9973)^3$=0.9919
P(No signal in 6 samples)=$(0.9973)^6$= 0.9839
P(No signal in 10 samples)=$(0.9973)^{10}$= 0.9733

15-111 Suppose that a process has a *PCR* = 2, but the mean is exactly 3 standard deviations above the upper specification limit. What is the probability of making a product outside the specification limits?

PCR = 2 but $\mu = \text{USL} + 3\sigma$

$$P(X < \text{USL}) = P\left(Z < \frac{(\mu - 3\sigma) - \mu}{\sigma}\right) = P(Z < -3) = 0.00135$$

15-112 A process is controlled by a *P* chart using samples of size 100. The center line on the chart is 0.05.

(a) What is the probability that the control chart detects a shift to 0.08 on the first sample following the shift?
(b) What is the probability that the control chart does not detect a shift to 0.08 on the first sample following the shift, but does detect it on the second sample?
(c) Suppose that instead of a shift in the mean to 0.08, the mean shifts to 0.10. Repeat parts (a) and (b).
(d) Compare your answers for a shift to 0.08 and for a shift to 0.10. Explain why they differ. Also, explain why a shift to 0.10 is easier to detect.

(a) The P(LCL < $\hat{P}$ < UCL), when p = 0.08, is needed.

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.05 - 3\sqrt{\frac{0.05(1-0.05)}{100}} = -0.015 \to 0$$

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.05 + 3\sqrt{\frac{0.05(1-0.05)}{100}} = 0.115$$

Therefore, when p = 0.08

$$P(0 \le \hat{P} \le 0.115) = P(\hat{P} \le 0.115) = P\left(\frac{\hat{P} - 0.08}{\sqrt{\frac{0.08(0.92)}{100}}} \le \frac{0.115 - 0.08}{\sqrt{\frac{0.08(0.92)}{100}}}\right)$$

$$= P(Z \le 1.29) = 0.90$$

Using the normal approximation to the distribution of $\hat{P}$. Therefore, the probability of detecting the shift on the first sample following the shift is 1 − 0.90 = 0.10.

(b) The probability that the control chart detects a shift to 0.08 on the second sample, but not the first, is (0.90)0.10= 0.09. This uses the fact that the samples are independent to multiply the probabilities.

(c) p = 0.10

$$P(0 \le \hat{P} \le 0.115) = P(\hat{P} \le 0.115) = P\left(\frac{\hat{P} - 0.10}{\sqrt{\frac{0.10(0.90)}{100}}} \le \frac{0.115 - 0.10}{\sqrt{\frac{0.10(0.90)}{100}}} \right)$$

$$= P(Z \le 0.5) = 0.69146$$

from the normal approximation to the distribution of $\hat{P}$. Therefore, the probability of detecting the shifton the first sample following the shift is 1 – 0.69146 = 0.30854.

The probability that the control chart detects a shift to 0.10 on the second sample after the shift, but not the first, is 0.69146(0.30854) =0.2133.

(d) A larger shift is generally easier to detect. Therefore, we should expect a shift to 0.10 to be detected quicker than a shift to 0.08.

15-113 Suppose that the average number of defects in a unit is known to be 8. If the mean number of defects in a unit shifts to 16, what is the probability that it is detected by a *U* chart on the first sample following the shift

(a) if the sample size is $n = 4$?
(b) if the sample size is $n = 10$?

Use a normal approximation for *U*.

$\bar{u} = 8$
(a) $n = 4$

$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} = 8 + 3\sqrt{\frac{8}{4}} = 12.24$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} = 8 - 3\sqrt{\frac{8}{4}} = 3.76$$

P($\bar{U}$ > 12.24 when λ = 16) = $P\left(Z > \frac{12.24 - 16}{\sqrt{16/4}} \right)$

= P(Z >–1.88) = 1 – P(Z <–1.88)= 1 – 0.03005= 0.96995

P($\bar{U}$ < 3.78) = $P\left(Z < \frac{3.76 - 16}{\sqrt{16/4}} \right)$

= P(Z <–6.12)
= 0

So the probability is 0.96995.

(b) $n = 10$

$$\text{UCL} = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} = 8 + 3\sqrt{\frac{8}{10}} = 10.68$$

$$\text{LCL} = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} = 8 - 3\sqrt{\frac{8}{10}} = 5.32$$

P(U > 10.68 when λ = 16) = $P\left(Z > \frac{10.68 - 16}{\sqrt{\frac{16}{10}}} \right)$ = P(Z >–4.22)= 1

So the probability is 1.

15-114 Suppose that the average number of defects in a unit is known to be 10. If the mean number of defects in a unit shifts to 14, what is the probability that it is detected by a *U* chart on the first sample following the shift

(a) if the sample size is $n = 1$?
(b) if the sample size is $n = 4$?

Use a normal approximation for *U*.

$\bar{u} = 10$
(a) $n = 1$

$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} = 10 + 3\sqrt{\frac{10}{1}} = 19.49$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} = 10 - 3\sqrt{\frac{10}{1}} = 0.51$$

P($\overline{U}$ > 19. 94 when λ = 14) = $P\left(Z > \frac{19.94 - 14}{\sqrt{14}}\right)$

= P(Z > 1.47)= 1 – P(Z < 1.47)= 1 – 0.9292= 0.0708

and

P($\overline{U}$ < 0.51) = $P\left(Z < \frac{0.51 - 14}{\sqrt{14}}\right) = 0$

(b) $n = 4$

$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}} = 10 + 3\sqrt{\frac{10}{4}} = 14.74$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}} = 10 - 3\sqrt{\frac{10}{4}} = 5.26$$

P($\overline{U}$ > 14.74 when λ = 14) = $P\left(Z > \frac{14.74 - 14}{\sqrt{\frac{14}{4}}}\right)$ = P(Z >0.40)= 1 – 0.6554= 0.3446

P($\overline{U}$ < 5.26 when λ = 14) = $P\left(Z < \frac{5.26 - 14}{\sqrt{\frac{14}{4}}}\right)$ = P(Z <–4.67) = 0

15-115 An EWMA chart with $\lambda = 0.5$ and $L = 3.07$ is to be used to monitor a process. Suppose that the process mean is $\mu_0 = 10$ and $\sigma = 2$.

(a) Assume that $n = 1$. What is the ARL without any shift in the process mean? What is the ARL to detect a shift to $\mu = 12$.
(b) Assume that $n = 4$. Repeat part (a) and comment on your conclusions.

(a) According to Table 16-10, if there is no shift, ARL=500.
If the shift in mean is $1\sigma_{\overline{X}}$, ARL=17.5.
(b) According to Table 16-10, if there is no shift, ARL=500.
If the shift in mean is $2\sigma_{\overline{X}}$, ARL=3.63.

15-116 The following table provides the costs for gasoline by month in the United States over recent years and the percentage of the cost due to refining, distribution and marketing, taxes, and crude oil. The table is from the U.S. Department of Energy Web site (http://tonto.eia.doe.gov/oog/info/gdu/gaspump.html). There is some concern that the refining or distribution and marketing percentages of the retail price have shown patterns over time.

(a) Construct separate control charts for the refining percentage of the retail price and the distribution and marketing percentage of the retail price. Use control charts for individual measurements. Comment on any signs of assignable causes on these charts.
(b) Construct a control chart for the crude oil percentage of the retail price. Use a control chart for individual measurements. Comment on any signs of assignable causes on this chart.
(c) Another way to study the data is to calculate refining, distribution and marketing, and tax as costs directly. The costs of these categories might not depend strongly on the crude oil cost. Use the percentages provided in the table to calculate the cost associated with refining and distribution and marketing each month. Construct separate control charts for refining and distribution and marketing costs each month. Use control charts for individual measurements. Comment on any signs of assignable causes on these charts and comment on any differences between these charts and the ones constructed in part (a).

What We Pay for in a Gallon of Regular Gasoline

Mo/Year	Retail Price (dollars per gallon)	Refining (%)	Distribution & Marketing (%)	Taxes (%)	Crude Oil (%)
Jan-00	1.289	7.8	13.0	32.1	47.1
Feb-00	1.377	17.9	7.5	30.1	44.6
Mar-00	1.517	15.4	12.8	27.3	44.6
Apr-00	1.465	10.1	20.2	28.3	41.4
May-00	1.485	20.2	9.2	27.9	42.7
Jun-00	1.633	22.2	8.8	25.8	43.1
Jul-00	1.551	13.2	15.8	27.2	43.8
Aug-00	1.465	15.8	7.5	28.8	47.8
Sep-00	1.550	15.4	9.0	27.2	48.3
Oct-00	1.532	13.7	10.1	27.5	48.6
Nov-00	1.517	10.4	11.8	27.8	50.0
Dec-00	1.443	8.0	17.9	29.2	44.8
Jan-01	1.447	17.8	10.4	29.2	42.7
Feb-01	1.450	17.3	11.0	29.1	42.6
Mar-01	1.409	18.8	9.7	30.0	41.5
Apr-01	1.552	31.6	4.6	27.1	36.7
May-01	1.702	26.4	14.0	24.7	35.0
Jun-01	1.616	13.2	24.1	26.0	36.7
Jul-01	1.421	10.0	20.0	30.0	40.0
Aug-01	1.421	20.0	9.0	30.0	41.0
Sep-01	1.522	18.0	17.0	28.0	37.0
Oct-01	1.315	10.0	20.8	31.9	37.2
Nov-01	1.171	10.0	18.0	36.0	36.0
Dec-01	1.086	11.7	12.7	38.7	36.9
Jan-02	1.107	13.0	11.8	37.9	37.2
Feb-02	1.114	12.1	11.2	37.7	39.1
Mar-02	1.249	19.4	6.1	33.6	40.9
Apr-02	1.397	15.5	13.0	30.1	41.4
May-02	1.392	11.9	14.2	30.2	43.7
Jun-02	1.382	15.0	13.0	30.4	41.6
Jul-02	1.397	15.0	12.6	30.1	42.3
Aug-02	1.396	11.4	13.4	30.0	45.0
Sep-02	1.400	10.8	12.6	30.0	46.7

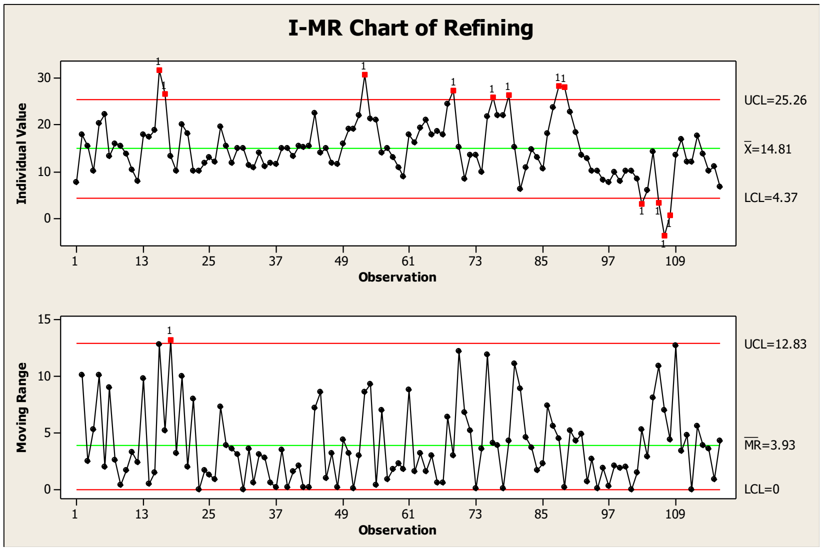
What We Pay for in a Gallon of Regular Gasoline					
Mo/Year	**Retail Price (dollars per gallon)**	**Refining (%)**	**Distribution & Marketing (%)**	**Taxes (%)**	**Crude Oil (%)**
Oct-02	1.445	13.9	11.7	29.1	45.3
Nov-02	1.419	11.1	18.0	29.6	41.3
Dec-02	1.386	11.7	12.3	30.3	45.7
Jan-03	1.458	11.5	10.3	28.8	49.4
Feb-03	1.613	15.0	9.5	26.0	49.5
Mar-03	1.693	14.8	14.8	24.8	45.5
Apr-03	1.589	13.2	19.8	26.4	40.5
May-03	1.497	15.3	16.3	28.1	40.4
Jun-03	1.493	15.1	12.3	28.1	44.5
Jul-03	1.513	15.3	11.9	27.8	44.9
Aug-03	1.620	22.5	8.2	25.9	43.3
Sept-03	1.679	13.9	22.7	25.0	38.3
Oct-03	1.564	14.9	16.1	26.9	42.2
Nov-03	1.512	11.7	15.3	27.8	45.2
Dec-03	1.479	11.5	12.6	28.4	47.5
Jan-04	1.572	15.9	9.9	26.7	47.5
Feb-04	1.648	19.1	9.2	25.5	46.2
Mar-04	1.736	19.0	11.3	24.2	45.5
Apr-04	1.798	22.0	9.9	23.4	44.6
May-04	1.983	30.6	7.8	21.2	40.4
Jun-04	1.969	21.3	16.7	21.3	40.7
Jul-04	1.911	20.9	11.3	21.9	45.8
Aug-04	1.878	13.9	12.2	22.4	51.5
Sep-04	1.870	14.8	9.1	22.5	53.6
Oct-04	2.000	13.0	9.3	21.0	56.7
Nov-04	1.979	10.7	14.6	21.2	53.6
Dec-04	1.841	8.9	18.1	23.9	49.1
Jan-05	1.831	17.7	7.3	24.0	50.9
Feb-05	1.910	16.1	9.3	23.0	51.6
Mar-05	2.079	19.3	6.2	21.2	53.4
Apr-05	2.243	20.9	9.6	19.6	49.8
May-05	2.161	17.9	12.8	20.4	49.0
Jun-05	2.156	18.5	6.9	20.4	54.2
Jul-05	2.290	17.9	8.0	19.2	54.9
Aug-05	2.486	24.3	2.1	17.7	55.9
Sep-05	2.903	27.3	7.5	15.2	50.0
Oct-05	2.717	15.1	17.8	16.2	50.9
Nov-05	2.257	8.3	13.1	19.5	57.1
Dec-05	2.185	13.5	7.9	20.1	58.4
Jan-06	2.316	13.4	6.6	19.8	60.1
Feb-06	2.280	9.8	11.4	20.1	58.6
Mar-06	2.425	21.7	4.5	18.9	54.8

What We Pay for in a Gallon of Regular Gasoline					
Mo/Year	**Retail Price (dollars per gallon)**	**Refining (%)**	**Distribution & Marketing (%)**	**Taxes (%)**	**Crude Oil (%)**
Apr-06	2.742	25.8	3.1	16.7	54.2
May-06	2.907	21.9	8.8	15.8	53.4
Jun-06	2.885	22.0	7.9	15.9	54.1
Jul-06	2.981	26.3	6.3	15.4	52.0
Aug-06	2.952	15.2	13.5	15.9	55.4
Sep-06	2.555	6.3	18.8	18.3	56.7
Oct-06	2.245	10.9	10.6	20.8	57.7
Nov-06	2.229	14.6	7.5	20.4	57.5
Dec-06	2.313	12.9	9.4	19.7	58.0
Jan-07	2.240	10.6	15.2	20.3	53.9
Feb-07	2.278	18.0	5.8	20.0	56.3
Mar-07	2.563	23.6	8.5	15.5	52.3
Apr-07	2.845	28.1	7.6	14.0	50.3
May-07	3.146	27.9	13.3	12.7	46.1
Jun-07	3.056	22.7	13.7	13.0	50.5
Jul-07	2.965	18.4	11.4	13.4	56.8
Aug-07	2.786	13.5	11.8	14.3	60.4
Sep-07	2.803	12.8	8.6	14.2	64.3
Oct-07	2.803	10.1	8.1	14.2	67.6
Nov-07	3.080	10.0	8.7	13.0	68.3
Dec-07	3.018	8.1	10.5	13.2	68.1
Jan-08	3.043	7.8	11.1	13.1	67.9
Feb-08	3.028	9.9	7.2	13.2	69.7
Mar-08	3.244	8.0	7.9	12.3	71.8
Apr-08	3.458	10.0	5.8	11.5	72.7
May-08	3.766	10.0	4.7	10.6	74.7
Jun-08	4.054	8.5	6.8	9.8	74.8
Jul-08	4.062	3.2	11.2	9.8	75.8
Aug-08	3.779	6.1	10.2	10.6	73.1
Sep-08	3.703	14.2	8.2	10.8	66.8
Oct-08	3.051	3.3	25.0	13.1	58.6
Nov-08	2.147	–3.7	24.7	18.6	60.4
Dec-08	1.687	0.7	19.5	23.6	56.2
Jan-09	1.788	13.4	10.7	22.3	53.6
Feb-09	1.923	16.8	14.9	20.7	47.6
Mar-09	1.959	12.0	12.3	20.4	55.3
Apr-09	2.049	12.0	12.1	19.5	56.4
May-09	2.266	17.6	–3.9	19.5	66.8
Jun-09	2.631	13.7	10.3	15.1	60.9
Jul-09	2.527	10.1	14.0	15.8	60.1
Aug-09	2.616	11.0	9.7	15.4	63.9
Sep-09	2.554	6.7	13.5	15.7	64.0

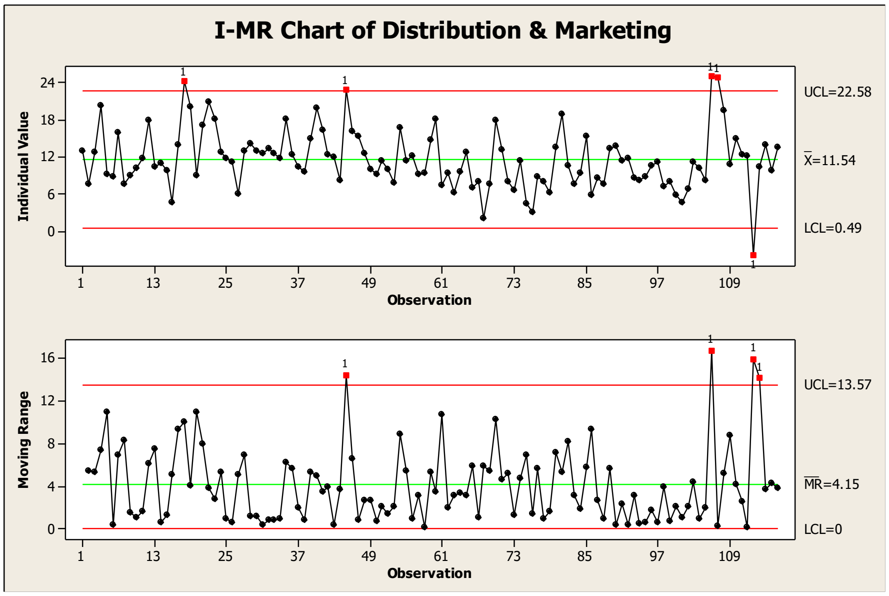
What We Pay for in a Gallon of Regular Gasoline					
Mo/Year	Retail Price (dollars per gallon)	Refining (%)	Distribution & Marketing (%)	Taxes (%)	Crude Oil (%)
Oct-09	2.551	6.5	9.1	15.7	68.6
Nov-09	2.651	4	11.7	15.1	69
Dec-09	2.607	6.6	11.2	15.4	66.6
Jan-10	2.715	5	10.9	14.7	69.1
Feb-10	2.644	5.9	9.8	15.1	69
Mar-10	2.772	9.3	7.5	14.4	68.5
Apr-10	2.848	10.4	6.9	14.1	68.4
May-10	2.836	8.4	13.3	14.2	63.9
Jun-10	2.732	9.4	10	14.7	65.6
Jul-10	2.729	9.4	9.5	14.8	66.1
Aug-10	2.73	5.7	11.9	14.8	67.4
Sep-10	2.705	6.5	11.3	14.9	67.1
Oct-10	2.801	5.3	9.7	14.4	70.5
Nov-10	2.859	4.4	10.2	14.1	71.1
Dec-10	2.993	9.9	8.4	13.5	68.1
Jan-11	3.095	10.6	9.1	13	67
Feb-11	3.211	14.2	7.9	12.5	65.2
Mar-11	3.561	13.1	7.2	11.3	68.3
Apr-11	3.8	15.7	4.8	10.6	68.7
May-11	3.906	14.1	10.2	10.3	65.2
Jun-11	3.68	11.5	10.9	10.9	66.5
Jul-11	3.65	14.8	6.3	11	67.7
Aug-11	3.639	15.9	10.4	11.1	62.5
Sep-11	3.611	14	12.2	11.1	62.5
Oct-11	3.448	10.9	8.7	11.7	68.5
Nov-11	3.384	−0.5	11.8	11.9	76.7
Dec-11	3.266	−1.5	9	12.3	80
Jan-12	3.38	6	6.4	11.9	75.5
Feb-12	3.579	11.8	5.2	11.2	71.5
Mar-12	3.852	15.7	5.9	10.8	67.4
Apr-12	3.9	14.6	8.4	10.6	66.1
May-12	3.732	13	10.3	11.1	65.4
Jun-12	3.539	11.8	14.4	11.7	61.9
Jul-12	3.439	14.9	7.9	12.1	64.9
Aug-12	3.722	17.5	7.1	11.2	64
Sep-12	3.849	17.9	8.8	10.8	62.2
Oct-12	3.746	12	12.8	11.1	64
Nov-12	3.452	9.2	11.9	12.1	66.6
Dec-12	3.31	8	11.3	12.6	68
Jan-13	3.319	8.5	6.9	12.6	71.8
Feb-13	3.67	15	6.8	11.4	66.7

Note that the original data contained two negative percentages and those were not changed.

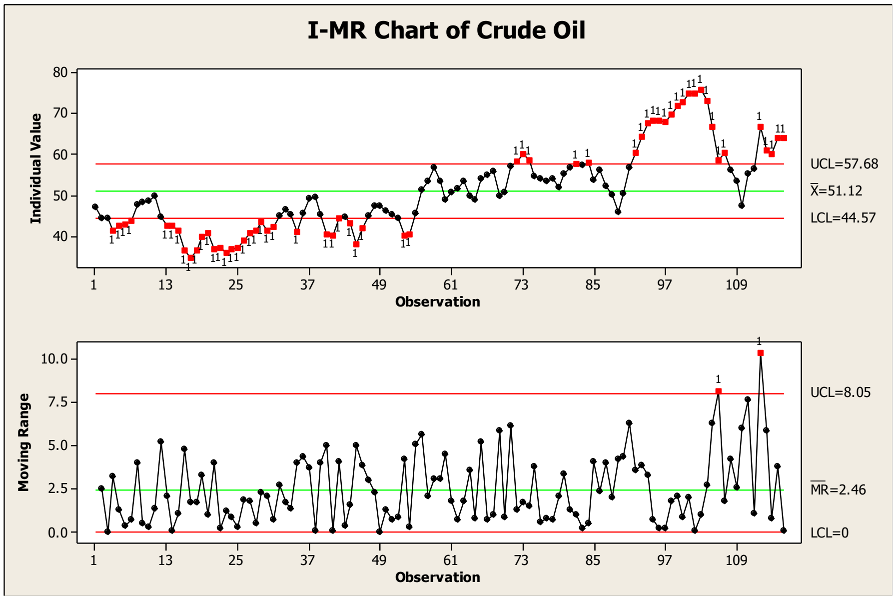
(a)I-MR chart for Refining percentage. The individual control chart shows that the process is out-of-control and observations 16, 17, 53, 69, 76, 79, 88, 89, 103, 106, 107, and 108 are beyond the control limits.



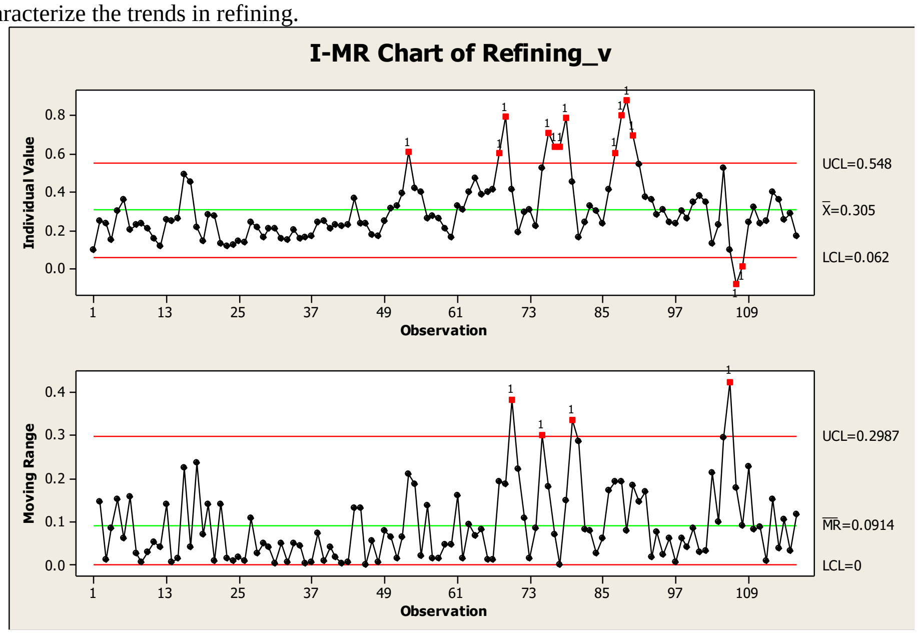
I-MR chart for Dist& Marketing percentage. The individual control chart shows that observations 18, 45, 106, 107, and 113 are beyond the control limits.



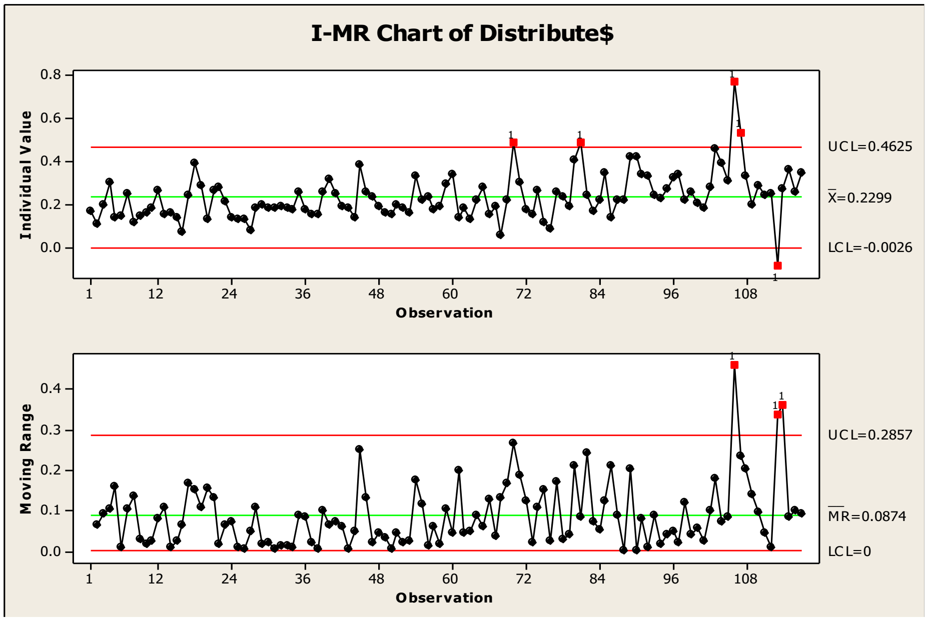
(b) I-MR chart for crude oil percentage of retail price. It appears there have been a multiple shifts in the process.



(c) The price is largely determined by the cost of crude. Changes to crude prices can change the percentage of price attributed to refining even when the refining costs are constant. The I-MR chart for Refining Cost follows. The individual control chart shows that observations 53, 68, 69, 76, 77, 78, 79, 87, 88, 89, 90,107, and 108 are beyond the control limits. It is similar to the I-MR from part a), but the new chart has fewer signals. One might consider this chart to better characterize the trends in refining.



The I-MR chart for Dist& Marketing costs follows. The individual control chart shows that observations 70, 81, 106, 107, and 113 are beyond the control limits. It is similar to the I-MR from part a), but the new chart does not signal at points earlier in time (at observations 18 and 45). One might consider this chart to better characterize the trends in marketing and distribution.



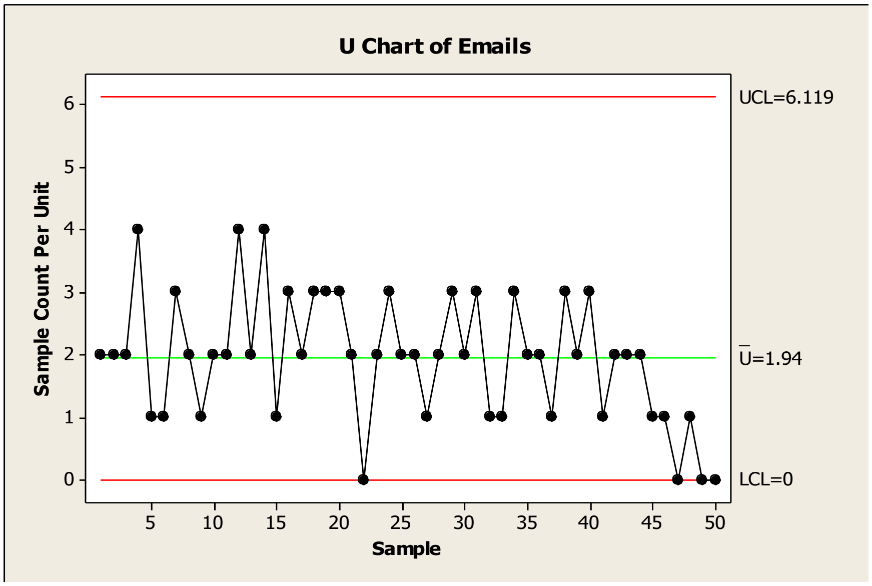
15-117 The following table shows the number of e-mails a student received each hour from 8:00 A.M. to 6:00 P.M. The samples are collected for five days from Monday to Friday.

Hour	M	T	W	Th	F
1	2	2	2	3	1
2	2	4	0	1	2
3	2	2	2	1	2
4	4	4	3	3	2
5	1	1	2	2	1
6	1	3	2	2	1
7	3	2	1	1	0
8	2	3	2	3	1
9	1	3	3	2	0
10	2	3	2	3	0

(a) Use the rational subgrouping principle to comment on why an $\overline{X}$ chart that plots one point each hour with a subgroup of size 5 is not appropriate.
(b) Construct an appropriate attribute control chart. Use all the data to find trial control limits, construct the chart, and plot the data.
(c) Use the trial control limits from part (b) to identify out-of-control points. If necessary, revise your control limits, assuming that any samples that plot outside the control limits can be eliminated.

(a) Differences between days of the week would change the variation in a subgroup. Therefore, the subgroup may contain sources of variation other than common cause variation.

(b) The data are ordered sequentially so that all hours in Monday are followed by all hours in Tuesday and so forth. In total there are 50 points plotted on a U chart with a subgroup size of one unit (one hour). The chart follows.



The points do not violate the control limits, but there is a downward trend in the last data points (end of the day on Friday).

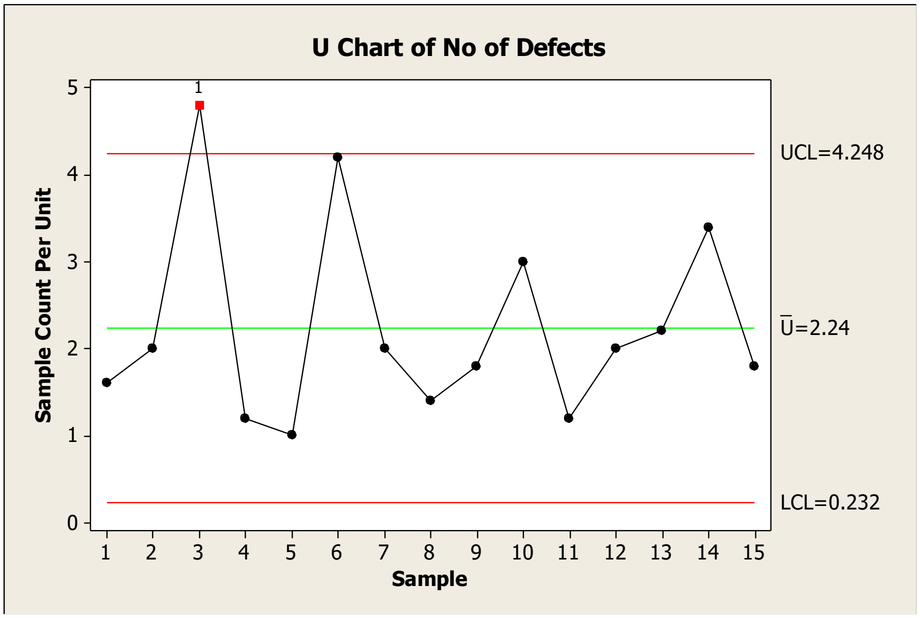
(c) Points do not violate the control limits in part (b) so the limits are not revised. However, the points in the downward trend might be considered to represent an assignable cause and these points might be eliminated. Then the limits might be revised.

15-118 The following are the number of defects observed on 15 samples of transmission units in an automotive manufacturing company. Each lot contains five transmission units.

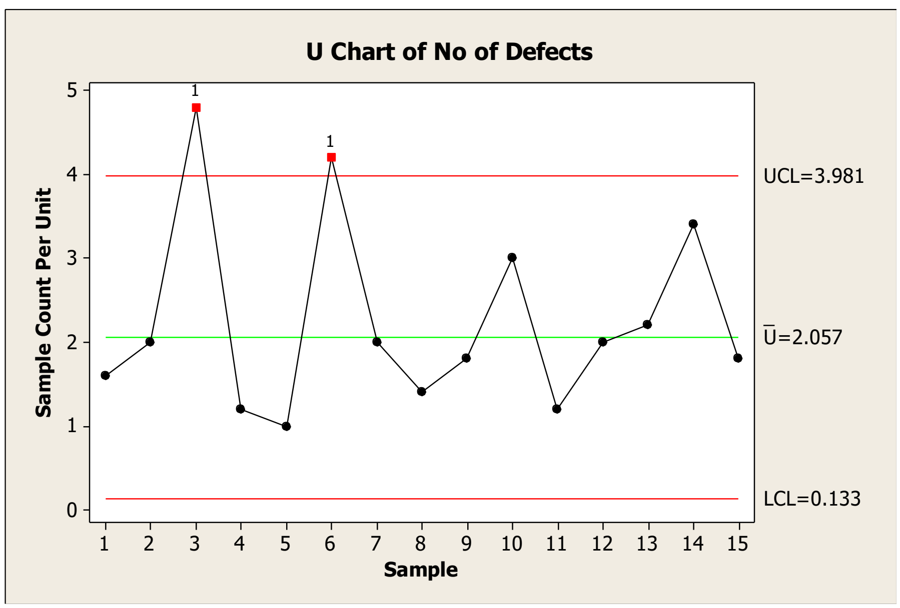
(a) Using all the data, compute trial control limits for a *U* control chart, construct the chart, and plot the data.
(b) Determine whether the process is in statistical control. If not, assume assignable causes can be found and out-of-control points eliminated. Revise the control limits.

Sample	No. of Defects	Sample	No. of Defects
1	8	11	6
2	10	12	10
3	24	13	11
4	6	14	17
5	5	15	9
6	21		
7	10		
8	7		
9	9		
10	15		

(a) The U chart follows.



(b) Point 3 violates the control limits in part (b). The control limits are revised one time by omitting the out-of-control point from the calculations (but the points are still plotted on the chart). However, an additional out-of-control signal at point 6 is shown.



15-119 Consider an $\overline{X}$ control chart with UCL = 32.802, UCL = 24.642, and n = 5. Suppose that the mean shifts to 30.

(a) What is the probability that this shift is detected on the next sample?
(b) What is the ARL to detect the shift?

(a) The difference $UCL - LCL = 6\hat{\sigma}_{\bar{X}} = 32.802 - 24.642 = 8.16$
Therefore, $\hat{\sigma}_{\bar{X}} = \frac{8.16}{6} = 1.36$ and $\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}}{\sqrt{5}}$. Therefore, $\hat{\sigma} = 1.36\sqrt{5} = 3.0411$
$P(24.642 < \bar{X} <$ 32.802) = P[(24.642 – 30)/1.36 < Z < (32.802 – 30)/1.36]
= P(-3.9397< Z < 2.0603) =0.9803.
Therefore the probability the shift is detected = 1 – 0.9803 = 0.0197 ≈ 0.02

(b) ARL = 1/*p*, where *p* is the probability a point exceeds a control limit. From part (b) *p* = 0.0197. Therefore, ARL = 1/0.0197 = 50.76

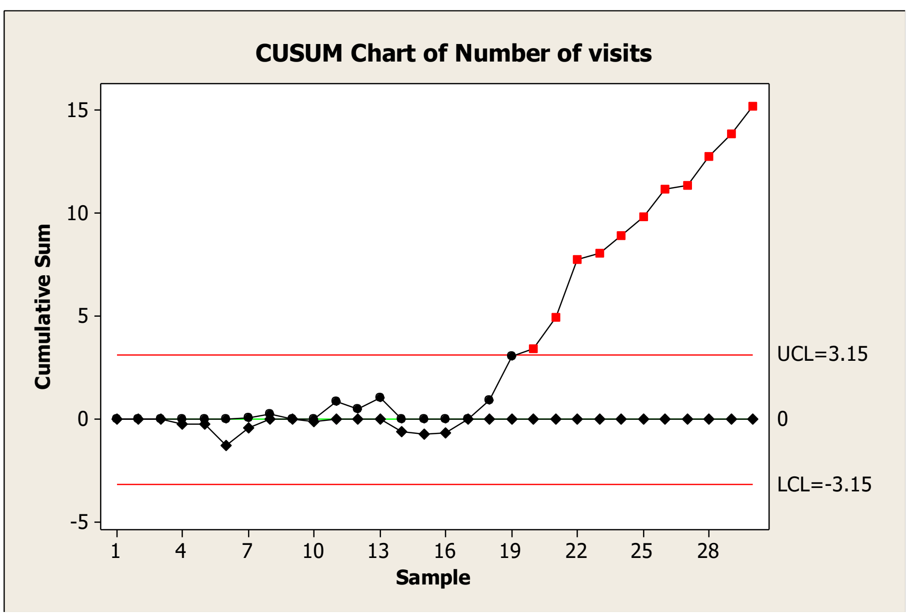
15-120 The number of visits (in millions) on a Web site is recorded every day. The following table shows the samples for 25 consecutive days.

(a) Estimate the process standard estimation.
(b) Set up a CUSUM control chart for this process, assuming the target is 10. Does the process appear to be in control?

Sample	Number of Visits	Sample	Number of Visits
1	10.12	16	9.66
2	9.92	17	10.42
3	9.76	18	11.30
4	9.35	19	12.53
5	9.60	20	10.76
6	8.60	21	11.92
7	10.46	22	13.24
8	10.58	23	10.64
9	9.95	24	11.31
10	9.50	25	11.26
11	11.26	26	11.79
12	10.02	27	10.53
13	10.95	28	11.82
14	8.99	29	11.47
15	9.50	30	11.76

(a) Process standard deviation is estimated using the average moving range of size 2 with MR/d2, where d2 = 1.128 for a moving range of 2. The estimate is 0.7887.

(b) A CUSUM chart with *h* = 4 and *k* = 0.5 follows. The process is not in control. A signal occurred at observation 20.

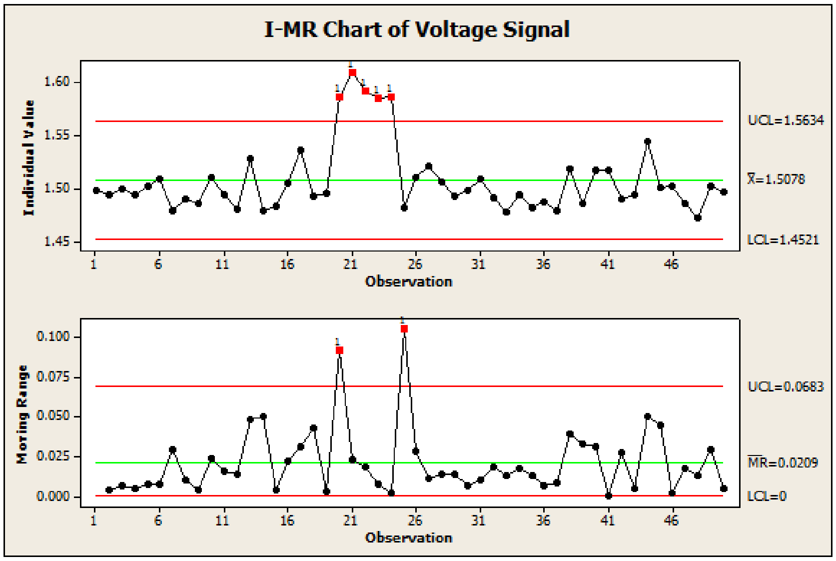


15-121 An article in *Microelectronics Reliability* ["Advanced Electronic Prognostics through System Telemetry and Pattern Recognition Methods," (2007, 47(12), pp. 1865–1873)] presented an example of electronic prognostics (a technique to detect faults in order to decrease the system downtime and the number of unplanned repairs in high-reliability and high-availability systems). Voltage signals from enterprise servers were monitored over time. The measurements are provided in the following table.

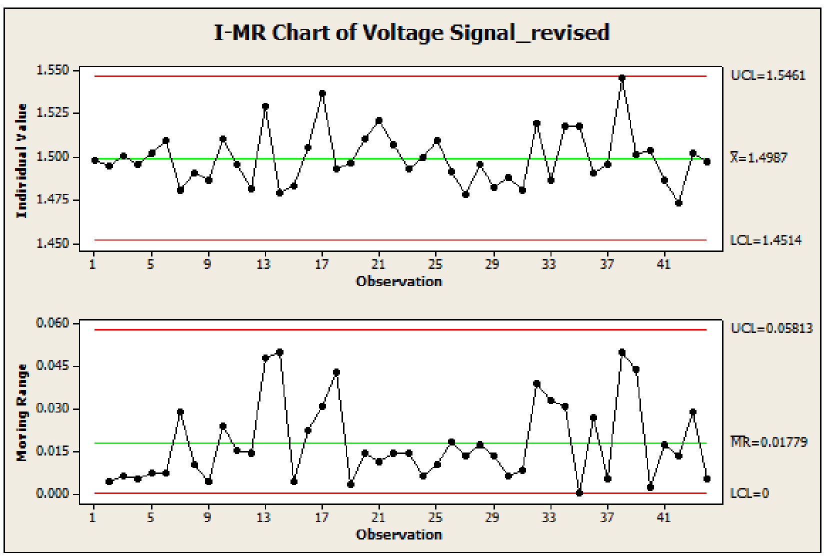
Observation	Voltage Signal	Observation	Voltage Signal
1	1.498	26	1.510
2	1.494	27	1.521
3	1.500	28	1.507
4	1.495	29	1.493
5	1.502	30	1.499
6	1.509	31	1.509
7	1.480	32	1.491
8	1.490	33	1.478
9	1.486	34	1.495
10	1.510	35	1.482
11	1.495	36	1.488
12	1.481	37	1.480
13	1.529	38	1.519
14	1.479	39	1.486
15	1.483	40	1.517
16	1.505	41	1.517
17	1.536	42	1.490
18	1.493	43	1.495
19	1.496	44	1.545
20	1.587	45	1.501
21	1.610	46	1.503
22	1.592	47	1.486
23	1.585	48	1.473
24	1.587	49	1.502
25	1.482	50	1.497

(a) Using all the data, compute trial control limits for individual observations and moving-range charts. Construct the chart and plot the data. Determine whether the process is in statistical control. If not, assume that assignable causes can be found to eliminate these samples and revise the control limits.
(b) Estimate the process mean and standard deviation for the in-control process.
(c) The report in the article assumed that the signal is normally distributed with a mean of 1.5 V and a standard deviation of 0.02 V. Do your results in part (b) support this assumption?

(a)



Both charts have points outside of the control limits. For the I chart, observations 20-24 are out of control. For the MR chart, observations 20 and 25 are out of control. Therefore, remove observations 20-25.



(b) $\hat{\mu} = \bar{x} = 1.499V$ and $\hat{\sigma} = \frac{\overline{MR}}{d_2} = \frac{0.01779}{1.128} = 0.016V$

(c) The results from the control charts are similar to those in the publication.

15-122 A article in the *Journal of Quality in Clinical Practice* ["The Application of Statistical Process Control Charts to the Detection and Monitoring of Hospital-Acquired Infections," (2001, Vol. 21, pp. 112–117)] reported the use of SPC methods to monitor hospital-acquired infections. The authors applied Shewhart, CUSUM, and EWMA charts to the monitor ESBL Klebsiella pneumonia infections. The monthly number of infections from June 1994 to April 1998 are shown in the following table.

(a) What distribution might be expected for these data? What type of control chart might be appropriate?
(b) Construct the chart you selected in part (a).

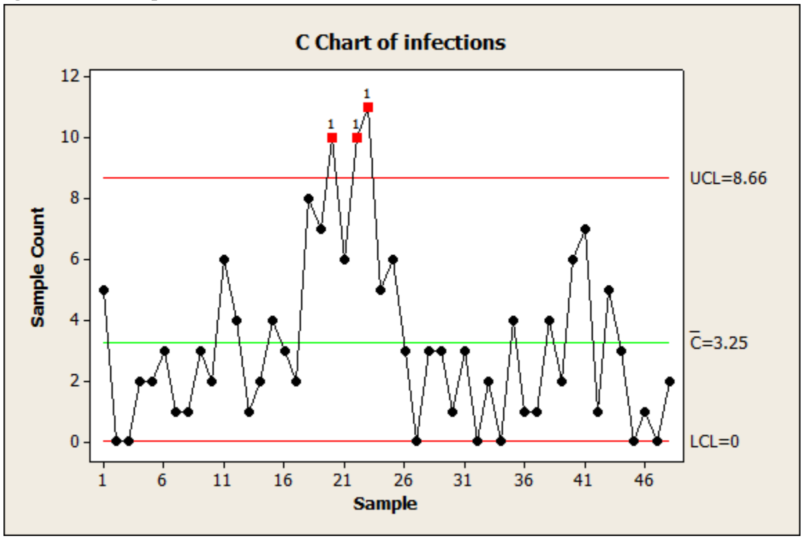
	Jan	Feb	Mar	April	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1994						5	0	0	2	2	3	1
1995	1	3	2	6	4	1	2	4	3	2	8	7
1996	10	6	10	11	5	6	3	0	3	3	1	3
1997	0	2	0	4	1	1	4	2	6	7	1	5
1998	3	0	1	0	2							

(c) Construct a CUSUM chart for these data with $k = 0\ 5$. and $h = 4$. The article included a similarly construced CUSUM chart. What is assumed for the distribution of the data in this chart? Can your CUSUM chart perform adequately?

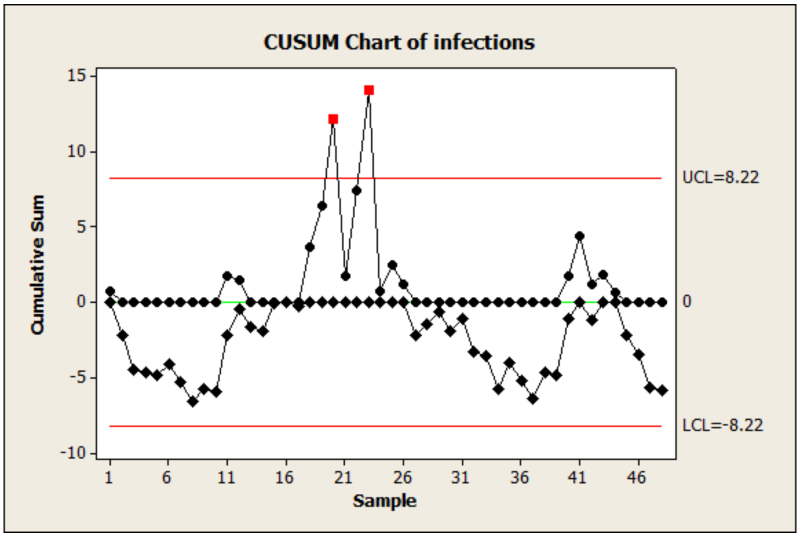
(d) Repeat part (c) for an EWMA chart with $\lambda = 0.2$.

(a) A Poisson distribution is expected, and a C chart might be appropriate for this data

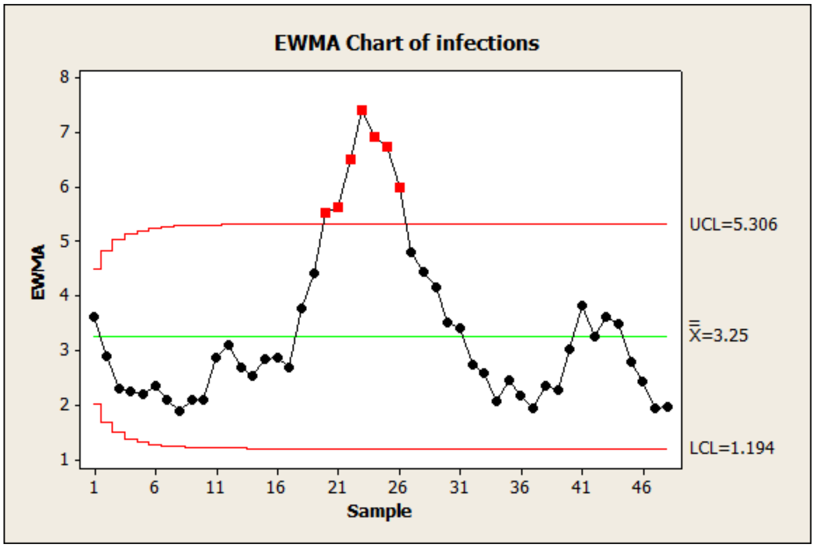
(b) For a C chart, signals occur at points 20,21,22



(c) For a CUSUM chart, signals occur at points 20 and 22. The CUSUM is reset after each signal.
The data is assumed to be normally distributed and for a Poisson distribution with larger values the normal approximation is reasonable. The CUSUM provides similar conclusions as the C chart.



(d) EWMA with lambda=0.2 Signals occur at points 20 – 26.



15-123 The control limits for an $\overline{X}$ chart with n = 4 are 12.8 and 24.8, and the *PCR* for a measurement is 1.33.

(a) Estimate the process standard deviation σ.
(b) Calculate the specification limits. Assume that they are centered around the process mean.

(a) 24-8 – 12.8 = 12 = $6\hat{\sigma}_{\bar{X}}$. Therefore, $\hat{\sigma}_{\bar{X}}=2=\frac{\hat{\sigma}_X}{\sqrt{n}}$. Therefore, $\hat{\sigma}_X=4$

(b) $PCR = \frac{USL-LSL}{6\cdot\sigma_X} = 1.33, USL - LSL = 31.92$. The half width is 31.92/2 = 15.96.
The process mean is centered between the control limits at 18.8.
Therefore, $USL = 18.8 + 15.96 = 34.76, LSL = 18.8 - 15.96 = 2.84$

15-124 Consider the turnaround time (TAT) for complete blood counts in Exercise 15-18. Suppose that the specifications for TAT are set at 20 and 80 minutes. Use the control chart summary statistics for the following.

(a) Estimate the process standard deviation.
(b) Calculate *PCR* and PCR_k for the process.

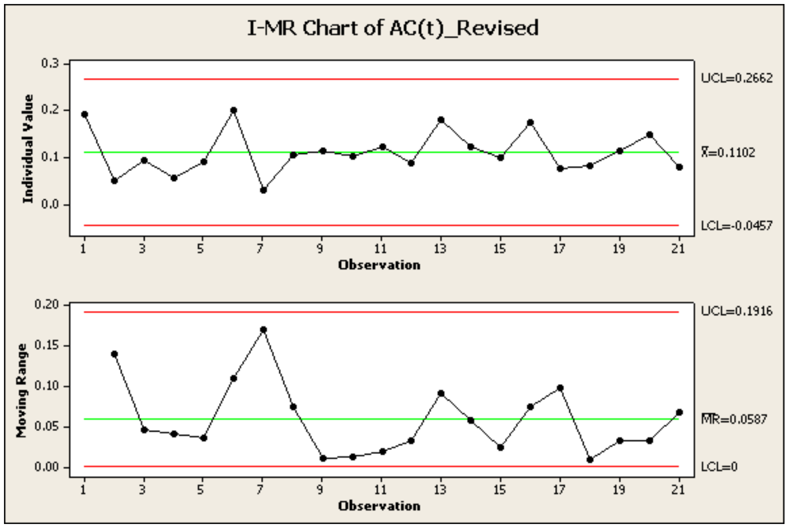
(a)$\bar{x} = 45.21, \bar{s} = 21$ Therefore, $\hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{21}{0.8862} = 23.70$

(b)

$$PCR = \frac{USL - LSL}{6 \cdot \sigma_X} = \frac{60}{6(23.70)} = 0.42$$

$$PCR_k = \min\left(\frac{80 - 45.21}{3(21)}, \frac{45.21 - 20}{3(21)}\right) = \min(0.49{,}0.35) = 0.35$$

15-125 Consider the inventory accuracy in Exercise 15-27. Because lower values are better, only the *UCL* = 0.3. is specified. Use the revised control chart to calculate PCR_k.



The revised control charts are shown.

$$\hat{\sigma} = \frac{\overline{MR}}{d_2} = \frac{0.0587}{1.128} = 0.052$$

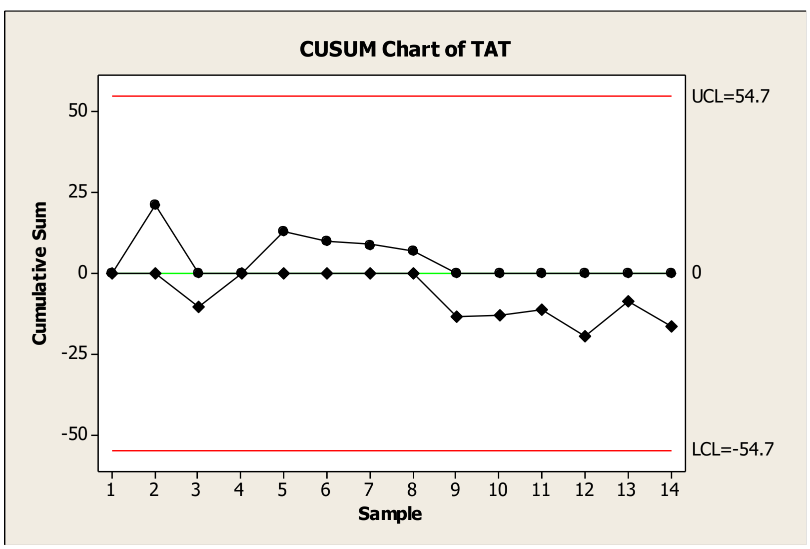
$$PCR_k = \left(\frac{USL - \bar{x}}{3\sigma_X}\right) = \frac{0.3 - 0.1102}{3(0.052)} = 1.22$$

15-126 Consider the TAT data in Exercise 15-18.

(a) Construct an CUSUM control chart with the target equal to the estimated process mean, *k* = 0.5, and *h* = 4. Does the process appear to be in control at the target?
(b) If the mean increases by 5 minutes, approximate the chart's ARL.

(a)$\bar{x} = 45.21, \bar{s} = 21$ Therefore, $\hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{21}{0.8862} = 23.70$and n = 3.

Therefore, $\hat{\sigma}_{\bar{X}} = \frac{\hat{\sigma}_X}{\sqrt{n}} = \frac{23.70}{\sqrt{3}} = 13.681$. Therefore K = 0.5(13.681) = 6.841 and H = 4(13.681) = 54.725.



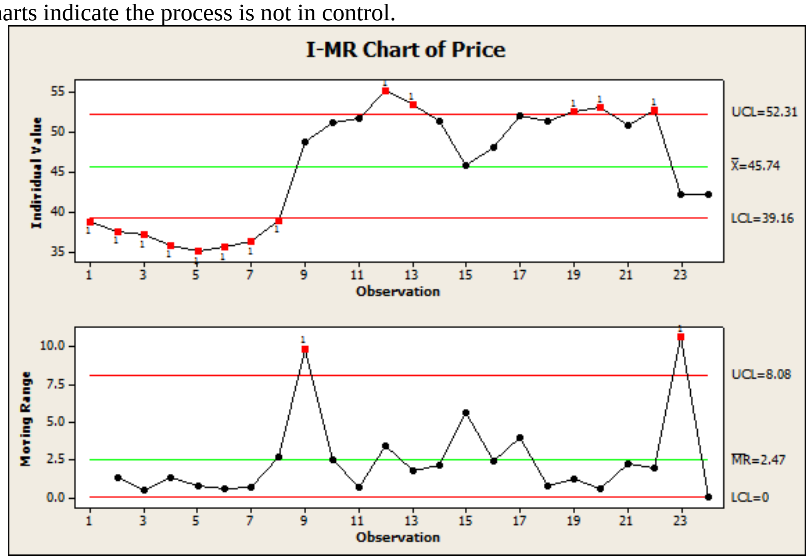
(b) If mean changes 5 minutes, then $\frac{\Delta}{\sigma} = \frac{5}{13.681} = 0.365$ and from the table the ARL is between 26.6 and 74.2.

15-127 An article in *Electric Power Systems Research* ["On the Self-Scheduling of a Power Producer in Uncertain Trading Environments" (2008, 78(3), pp. 311–317)] considered a self-scheduling approach for a power producer. The following table shows the forecasted prices of energy for a 24-hour time period according to a base case scenario.

Hour	Price	Hour	Price	Hour	Price
1	38.77	9	48.75	17	52.07
2	37.52	10	51.18	18	51.34
3	37.07	11	51.79	19	52.55
4	35.82	12	55.22	20	53.11
5	35.04	13	53.48	21	50.88
6	35.57	14	51.34	22	52.78
7	36.23	15	45.8	23	42.16
8	38.93	16	48.14	24	42.16

(a) Construct individuals and moving-range charts. Determine whether the energy prices fluctuate in statistical control.
(b) Is the assumption of independent observations reasonable for these data?

(a) The control charts indicate the process is not in control.

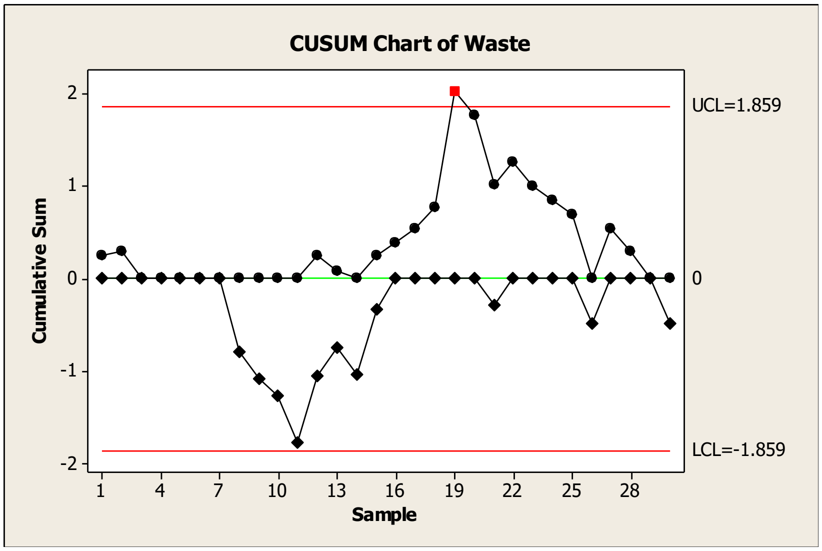


(b) No, the assumption of independent observation is not reasonable. The control chart does not perform as expected for non-independent data.

15-128 Consider the infectious-waste data in Exercise 15-26. Use the data after the process change only.

(a) Construct an CUSUM control chart with the target equal to the estimated process mean, $k = 0.5$, and $h = 4$. Does the process appear to be in control at the target?
(b) If the mean increases by 1.0 lb, approximate the ARL of the chart.

(a) The CUSUM chart uses the 30 observations after the change. The chart provides similar conclusions to those from the individuals chart. Observation 19 generates a signal on the CUSUM chart.



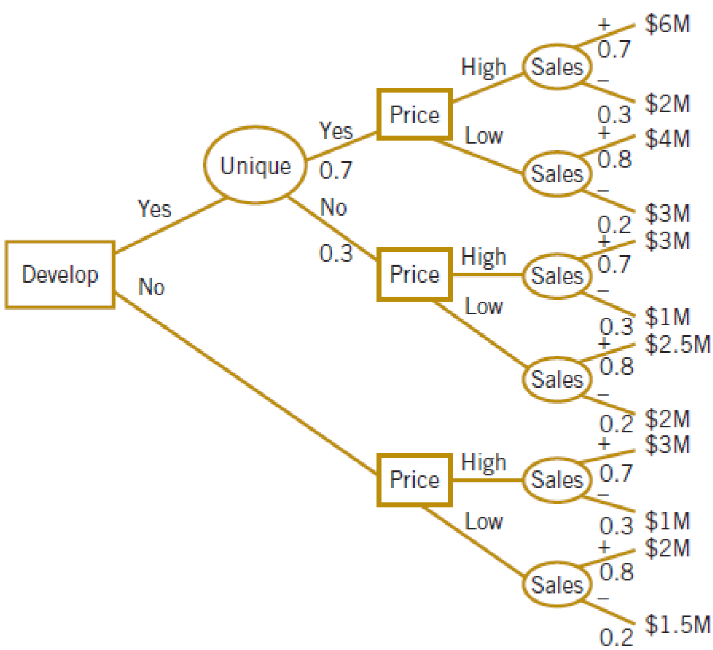
(b) From the individuals and moving range charts, $\bar{x} = 4.523, \hat{\sigma} = \frac{\overline{MR}}{d_2} = \frac{0.524}{1.128} = 0.465$
Therefore, K = 4(0.465) = 1.86 and H= 0.5(0.418) = 0.23

If the mean shifts by 1.0, this equals 1/0.465 = 2.15σ . The ARL for a shift of 2σ is 3.34.

15-129 Reconsider the extended warranty decision in Example 15-8. Determine the cost of the extended warranty so that the expected costs of the actions to either purchase the warranty or not are equal.

From the example, the expected cost if an extended warranty is not purchased is $270. Therefore, if the warranty costs $270 the actions to purchase or not have the same expected cost.

15-130 Analyze the develop or contract decision problem in Example 15-9 based on the best-case (optimistic) criterion and determine the actions selected at each decision node. Do any actions differ from those selected in the example?



Decisions:
1. When a new product is developed and a unique product is achieved, the optimistic outcomes for the high and low prices are 6M and 4M, respectively. Therefore, the price is set high.

2. When a new product is developed and a unique product is not achieved, the optimistic outcomes for the high and low prices are 3M and 2.5M, respectively. Therefore, the price is set high.

3. When a new product is not developed, the optimistic outcomes for the high and low prices are 3M and 2M, respectively. Therefore, the price is set high.

4. When a new product is developed, and if a unique product is achieved, the optimistic action is to price high with an outcome of 6M. When a new product is developed, and if a unique product is not achieved, the optimistic action is to price high with an outcome of 3M. Therefore, the optimistic outcome from the unique node is 6M.

When a new product is not developed, the price decision based on the most optimistic outcome is to price high with the optimistic outcome 3M.

Therefore, the decision is to develop a new product. This differs from the example.

Mind Expanding Exercises

15-131 Suppose that a process is in control, and 3-sigma control limits are in use on an $\overline{X}$ chart. The subgroup size is 4. Let the mean shift by 1.5σ. What is the probability that this shift remains undetected for three consecutive samples? What would its probability be if 2-sigma control limits were used?

Let p denote the probability that a point plots outside of the control limits when the mean has shifted from μ_0 to $\mu = \mu_0 + 1.5\sigma$. Then:

$$
\begin{aligned}
1-p &= P\left(LCL < \overline{X} < UCL \mid \mu = \mu_0 + 1.5\sigma\right) \\
&= P\left(\mu_0 - 3\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu_0 + 3\frac{\sigma}{\sqrt{n}} \mid \mu = \mu_0 + 1.5\sigma\right) \\
&= P\left(\frac{-1.5\sigma}{\sigma/\sqrt{n}} - 3 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{-1.5\sigma}{\sigma/\sqrt{n}} + 3\right) \\
&= P\left(-3 - 1.5\sqrt{n} < Z < +3 - 1.5\sqrt{n}\right) \text{ when } n = 4 \\
&= P(-6 < Z < 0) = 0.5
\end{aligned}
$$

Therefore, the probability the shift is undetected for three consecutive samples is $(1-p)^3 = 0.5^3 = 0.125$.

If 2-sigma control limits were used, then

$$
\begin{aligned}
1-p &= P\left(LCL < \overline{X} < UCL \mid \mu = \mu_0 + 1.5\sigma\right) = P\left(\mu_0 - 2\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu_0 + 2\frac{\sigma}{\sqrt{n}} \mid \mu = \mu_0 + 1.5\sigma\right) \\
&= P\left(\frac{-1.5\sigma}{\sigma/\sqrt{n}} - 2 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{-1.5\sigma}{\sigma/\sqrt{n}} + 2\right) \text{ when } n = 4 \\
&= P(-5 < Z < -1) = 0.1587 - 0 = 0.1587
\end{aligned}
$$

Therefore, the probability the shift is undetected for three consecutive samples is $(1-p)^3 = 0.1587^3 = 0.004$.

15-132 Consider an $\overline{X}$ control chart with k-sigma control limits and subgroup size n. Develop a general expression for the probability that a point plots outside the control limits when the process mean has shifted by δ units from the center line.

$$
\begin{aligned}
LCL &= \mu_0 - k\sigma/\sqrt{n} \\
CL &= \mu_0 \\
UCL &= \mu_0 + k\sigma/\sqrt{n}
\end{aligned}
$$

(a)

$$1-p=1-P\left(LCL<\overline{X}<UCL \mid \mu=\mu_0+\delta\sigma\right)$$
$$=1-P\left(\mu_0-k\frac{\sigma}{\sqrt{n}}<\overline{X}<\mu_0+k\frac{\sigma}{\sqrt{n}} \mid \mu=\mu_0+\delta\sigma\right)$$
$$1=1-P\left(-k-\frac{\delta\sigma}{\sigma/\sqrt{n}}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<k-\frac{\delta\sigma}{\sigma/\sqrt{n}}\right)$$
$$=1-P(-k-\delta\sqrt{n}<Z<k-\delta\sqrt{n})$$
$$=1-[\Phi(k-\delta\sqrt{n})-\Phi(-k\delta\sqrt{n}]$$

where Φ(Z) is the standard normal cumulative distribution function.

15-133 Suppose that an $\overline{X}$ chart is used to control a normally distributed process and that samples of size *n* are taken every *n* hours and plotted on the chart, which has *k*-sigma limits.

(a) Find a general expression for the expected number of samples and time that is taken until a false signal is generated.
(b) Suppose that the process mean shifts to an out-of-control state, say $\mu_1 = \mu_0 + \delta\sigma$. Find an expression for the expected number of samples that is taken until a false action is generated.
(c) Evaluate the in-control ARL for *k* = 3. How does this change if *k* = 2? What do you think about the use of 2-sigma limits in practice?
(d) Evaluate the out-of-control ARL for a shift of 1 sigma, given that *n* = 5.

$$LCL=\mu_0-k\sigma/\sqrt{n}$$
$$CL=\mu_0$$
$$UCL=\mu_0+k\sigma/\sqrt{n}$$

(a) ARL = 1/*p* where *p* is the probability a point plots outside of the control limits. Then,

$$1-p=P\left(LCL<\overline{X}<UCL \mid \mu_0\right)=P\left(-k<\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}<k \mid \mu_0\right)$$
$$=P(-k<Z<k)=\Phi(k)-\Phi(-k)=2\Phi(k)-1$$

where Φ(Z) is the standard normal cumulative distribution function. Therefore, *p* = 2 – 2Φ(k) andARL=1/[2–2Φ(k)]. The mean time until a false alarm is 1/*p* hours.

(b) ARL=1/*p* where

$$1-p=P\left(LCL<\overline{X}<UCL \mid \mu_1=\mu_0+\delta\sigma\right)=P\left(-k-\frac{\delta\sigma}{\sigma/\sqrt{n}}<\overline{X}<k-\frac{\delta\sigma}{\sigma/\sqrt{n}}\right)$$
$$=P\left(-k-\sqrt{n}\delta<Z<k-\sqrt{n}\delta\right)$$
$$=\Phi(k-\sqrt{n}\delta)-\Phi(-k-\sqrt{n}\delta)$$

and

$$p=1-\Phi(k-\sqrt{n}\delta)+\Phi(-k-\sqrt{n}\delta)$$

(c) ARL=1/*p* where 1-*p* = P(-3<Z<3)= 0.9973. Thus, ARL = 1/0.0027=370.4.

If *k* = 2, 1-*p* = P(-2<Z<2)=0.9545 and ARL=1/*p* = 22.0.

The 2-sigma limits result in a false alarm for every 22 points on the average. This is a high number of false alarms for the routine use of a control chart.

(d) From part (b),ARL = 1/p, where 1-p = $P(-3-\sqrt{5} < Z < 3-\sqrt{5}) = 0.7764$. ARL = 4.47 assuming 3-sigmacontrol limits.

15-134 Suppose that a *P* chart with center line at $\overline{p}$ with *k*-sigma control limits is used to control a process. There is a critical fraction defective p_c that must be detected with probability 0.50 on the first sample following the shift to this state. Derive a general formula for the sample size that should be used on this chart.

Determine*n* such that

$$0.5 = P(LCL < \hat{P} < UCL \mid p = p_c)$$

$$= P\left(\frac{\overline{p} - k\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} - p_c}{\sqrt{\frac{p_c(1-p_c)}{n}}} < \frac{\overline{P} - p_c}{\sqrt{\frac{p_c(1-p_c)}{n}}} < \frac{\overline{p} + k\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} - p_c}{\sqrt{\frac{p_c(1-p_c)}{n}}} \mid p = p_c \right)$$

$$= P\left(\frac{\overline{p} - p_c}{\sqrt{\frac{p_c(1-p_c)}{n}}} - \frac{k\sqrt{\overline{p}(1-\overline{p})}}{\sqrt{p_c(1-p_c)}} < Z < \frac{\overline{p} - p_c}{\sqrt{\frac{p_c(1-p_c)}{n}}} - \frac{k\sqrt{\overline{p}(1-\overline{p})}}{\sqrt{p_c(1-p_c)}} \right)$$

Use the fact that if $p_c > \overline{p}$ then the probability is approximately equal to the probability that *Z* is less than the upper limit.

$$\cong P\left(Z < \frac{\overline{p} - p_c}{\sqrt{\frac{p_c(1-p_c)}{n}}} - \frac{k\sqrt{\overline{p}(1-\overline{p})}}{\sqrt{p_c(1-p_c)}} \right)$$

Then, the probability above approximately equals 0.5 if

$$\frac{\overline{p} - p_c}{\sqrt{\frac{p_c(1-p_c)}{n}}} = \frac{k\sqrt{\overline{p}(1-\overline{p})}}{\sqrt{p_c(1-p_c)}}$$

Solving for *n*, $n = \dfrac{k^2\overline{p}(1-\overline{p})}{(\overline{p} - p_c)^2}$

15-135 Suppose that a *P* chart with center line at $\overline{p}$ and *k*-sigma control limits is used to control a process. What is the smallest sample size that can be used on this control chart to ensure that the lower control limit is positive?

The LCL is $\overline{p} - k\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$

$\overline{p} - k\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$ =0 or $n = \dfrac{k^2(1-\overline{p})}{\overline{p}}$

15-136 A process is controlled by a *P* chart using samples of size 100. The center line on the chart is 0.05. What is the probability that the control chart detects a shift to 0.08 on the first sample following the shift? What is the probability that the shift is detected by at least the third sample following the shift?

The $P(LCL < \hat{P} < UCL \mid p = 0.08)$ is desired. Now, using the normal approximation:

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.05 - 3\sqrt{\frac{0.05(0.95)}{100}} = -0.015 \to 0$$

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.05 + 3\sqrt{\frac{0.05(0.95)}{100}} = 0.115$$

$$P(0 < \hat{P} < 0.115 \mid p = 0.08) = P(\hat{P} < 0.115 \mid p = 0.08)$$

$$= P\left(\frac{\hat{P} - 0.08}{\sqrt{\frac{0.08(1-0.08)}{100}}} < \frac{0.115 - 0.08}{\sqrt{\frac{0.08(1-0.08)}{100}}}\right) = P(Z < 1.29) = 0.90$$

Therefore, the probability of detecting shift on the first sample following the shift is 1 – 0.90 = 0.10.

The probability of detecting a shift by at least the third sample following the shift can be determined from the geometric distribution to be 0.10 + 0.90(0.10) +0.90^2(0.10) = 0.27

15-137 Consider a process whose specifications on a quality characteristic are 100 ± 15. You know that the standard deviation of this normally distributed quality characteristic is 5. Where should you center the process to minimize the fraction defective produced? Now suppose that the mean shifts to 105, and you are using a sample size of 4 on an $\bar{X}$ chart.

(a) What is the probability that such a shift is detected on the first sample following the shift?
(b) What is the average number of samples until an out-of-control point occurs? Compare this result to the average number of observations until a defective occurs (assuming normality).

The process should be centered at the middle of the specifications; that is, at 100.

For an $\bar{X}$ chart:

$$CL = \mu_0 = 100$$

$$LCL = \mu_0 - k\sigma/\sqrt{n} = 100 - 3(5)/2 = 92.5$$

$$UCL = \mu_0 + k\sigma/\sqrt{n} = 100 + 3(5)/2 = 107.5$$

$$P(LCL < \bar{X} < UCL \mid \mu = 105) = P\left(\frac{92.5-105}{5/2} < \frac{\bar{X}-105}{5/2} < \frac{107.5-105}{5/2}\right)$$

$$= P(-5 < Z < 1) = 0.84$$

The requested probability is then 1 – 0.84 = 0.16. The ARL = 1/0.16 = 6.25. With μ = 105, the specifications at 100±15 and σ = 5, the probability of a defective item is

$$P(X < 85) + P(X > 115) = P\left(\frac{X-105}{5} < \frac{85-105}{5}\right) + P\left(\frac{X-105}{5} > \frac{115-105}{5}\right)$$

$$= P(Z < -4) + P(Z > 2) = 0.0228$$

Therefore, the average number of observations until a defective occurs, follows from the geometric distribution to be 1/0.0228 = 43.86. However, the $\bar{X}$ chart only requires 6.25 samples of 4 observations each = 6.25(4) = 25 observations, on average, to detect the shift.

15-138 ***NP* Control Chart**. An alternative to the control chart for fraction defective is a control chart based on the number of defectives or the *NP* control chart. The chart has center line at $n\bar{p}$, the control limits are

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$$

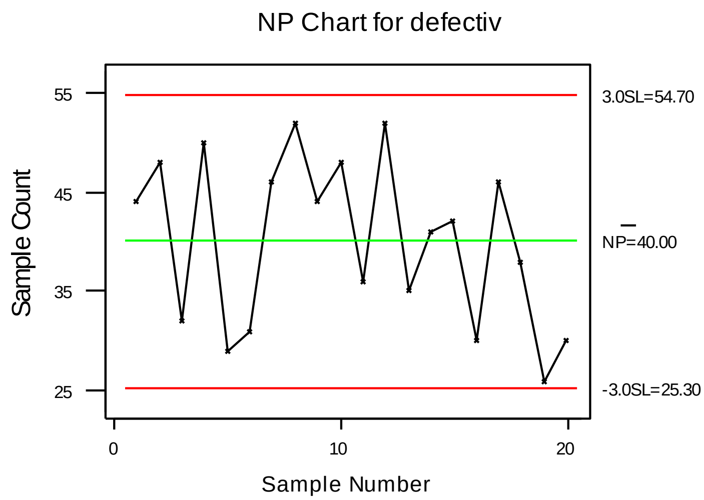
$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$$

and the number of defectives for each sample is plotted on the chart.

(a) Verify that the control limits provided are correct.
(b) Apply this control, chart to the data in Example 15-4.
(c) Will this chart always provide results that are equivalent to the usual *P* chart?

(a) Let *X* denote the number of defectives in a sample of *n*. Then *X* has a binomial distribution with E(*X*) = *np* and V(*X*) = *np*(1-*p*). Therefore, the estimates of the mean and standard deviation of *X* are $n\overline{p}$ and $\sqrt{n\overline{p}(1-\overline{p})}$, respectively. Using these estimates results in the given control limits.

(b) Data from example 16-4



(c)

$$n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} < n\hat{P} < n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})}$$

$$\overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} < \hat{P} < \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

Therefore, the *np* control chart always provides results equivalent to the *p* chart.

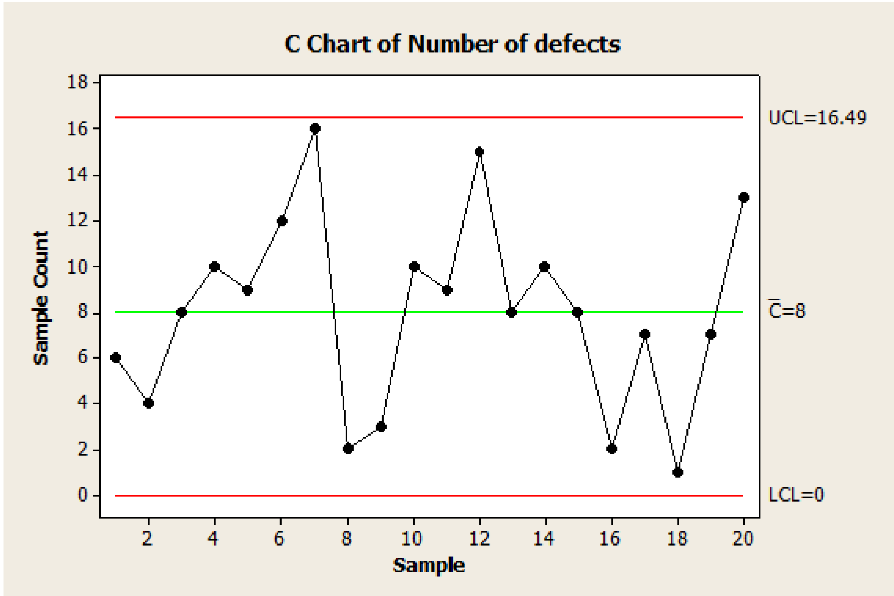
15-139 ***C* Control Chart**. An alternative to the *U* chart is a chart based on the number of defects. The chart has center line at $n\bar{u}$, and the control limits are

$$UCL = n\overline{u} + 3\sqrt{n\overline{u}}$$
$$LCL = n\overline{u} - 3\sqrt{n\overline{u}}$$

(a) Verify that the control limits provided are correct.
(b) Apply this chart to the data in Example 15-5.
(c) Will this chart always provide results equivalent to the *U* chart?

(a)

The center line $\overline{C} = 8$, and UCL=16.49 and LCL=0



(b)Yes

15-140 **Standardized Control Chart**. Consider the *P* chart with the usual 3-sigma control limits. Suppose that we define a new variable

$$Z_i = \frac{P_i - \overline{P}}{\sqrt{\dfrac{\overline{P}(1-\overline{P})}{n}}}$$

as the quantity to plot on a control chart. It is proposed that this new chart has a center line at 0 with the upper and lower control limits at ±3. Verify that this standardized control chart is equivalent to the original *P* chart.

Because $-3 < Z < 3$ if and only if

$$-3 < \frac{\hat{P} - \overline{p}}{\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}} < 3 \quad or \quad \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} < \hat{P}_i < \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

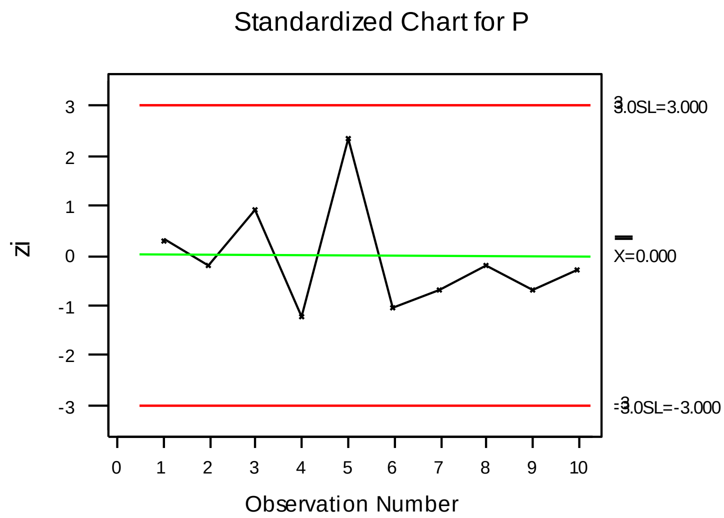
a point is in control on this chart if and only if the point is in control on the original *p* chart.

15-141 **Unequal Sample Sizes**. One application of the standardized control chart introduced in Exercise 15-140 is to allow unequal sample sizes on the control chart. Provide details concerning how this procedure would be implemented and illustrate using the following data:

Sample, i	1	2	3	4	5	6	7	8	9	10
n_i	20	25	20	25	50	30	25	25	25	20
p_i	0.2	0.16	0.25	0.08	0.3	0.1	0.12	0.16	0.12	0.15

For unequal sample sizes, the *p* control chart can be used with the value of *n* equal to the size of each sample. That is,

$$Z_i = \frac{\hat{P} - \overline{p}}{\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n_i}}}$$ where n_i is the size of the *i*th sample.


Standardized Chart for P
3.0SL=3.000
X=0.000
-3.0SL=-3.000
zi
3
2
1
0
-1
-2
-3
0 1 2 3 4 5 6 7 8 9 10
Observation Number